Published in final edited form as:

Toxicol Appl Pharmacol. 2015 January 1; 282(1): 30–41. doi:10.1016/j.taap.2014.10.017.

Analysis of the *AHR* gene proximal promoter GGGGC-repeat polymorphism in lung, breast, and colon cancer

Barbara C. Spink^a, Michael S. Bloom^b, Susan Wu^a, Stewart Sell^{a,c}, Erasmus Schneider^{a,c}, Xinxin Ding^{a,b,c}, and David C. Spink^{a,b,*}

^aWadsworth Center, New York State Department of Health, Albany, NY 12201, United States

^bDepartment of Environmental Health Sciences, School of Public Health, University at Albany, State University of New York, Albany, NY 12201, United States

^cDepartment of Biomedical Sciences, School of Public Health, University at Albany, State University of New York, Albany, NY 12201, United States

Abstract

The aryl hydrocarbon receptor (AhR) regulates expression of numerous genes, including those of the *CYP1* gene family. With the goal of determining factors that control *AHR* gene expression, our studies are focused on the role of the short tandem repeat polymorphism, (GGGGC)_n, located in the proximal promoter of the human *AHR* gene. When luciferase constructs containing varying GGGGC repeats were transfected into cancer cell lines derived from lung, colon, and breast, the number of GGGGC repeats affected *AHR* promoter activity. The number of GGGGC repeats was determined in DNA from 327 humans and from 38 samples representing 5 species of non-human primates. In chimpanzees and 3 species of macaques, only (GGGGC)₂ alleles were observed; however, in western gorilla, (GGGGC)_n alleles with n = 2, 4, 5, 6, 7, and 8 were identified. In all human populations examined, the frequency of (GGGGC)_n was n = 4 > 5 >> 2, 6. When frequencies of the (GGGGC)_n alleles in DNA from patients with lung, colon, or breast cancer were evaluated, the occurrence of (GGGGC)₂ was found to be 8-fold more frequent among lung cancer patients in comparison with its incidence in the general population, as represented by New York State neonates. Analysis of matched tumor and non-tumor DNA samples from the same individuals provided no evidence of microsatellite instability. These studies indicate that the (GGGGC)_n short tandem repeats are inherited, and that the (GGGGC)₂ allele in the *AHR* proximal promoter region should be further investigated with regard to its potential association with lung cancer susceptibility.

© 2014 Elsevier Inc. All rights reserved.

*Corresponding author. Tel.: +1 518 486 2530; fax: +1 518 473 2895. spink@wadsworth.org.

Conflict of interest statement: The authors have nothing to disclose.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

aryl hydrocarbon receptor; (GGGGC)_n repeat polymorphism; short tandem repeat; lung cancer; breast cancer; colon cancer

1. Introduction

Since its discovery as the receptor mediating the effects of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) and polycyclic aromatic hydrocarbons (PAHs) (Nebert and Gelboin, 1968; Poland et al., 1976; Conney, 1982; 2003; Oakey, 2007; Xing et al., 2012), the aryl hydrocarbon receptor (AhR) has been identified as a key regulatory factor of carcinogenesis in the lung and other tissues. The AhR transcriptionally regulates expression of the cytochrome P450 family 1 (CYP1) enzymes that metabolize PAHs, heterocyclic aromatic amines and other exogenous and endogenous compounds. The AhR-mediated induction of the CYP1 enzymes that catalyze the metabolic activation of PAHs to DNA adductive forms is integral to the carcinogenic process (Nebert et al., 2000; 2004), as is evidenced by the fact that AhR-null mice are refractory to PAH-induced carcinogenesis (Shimizu et al., 2000). However, induction of phase I enzymes by exposure to agents including TCDD can in some instances inhibit carcinogenicity (DiGiovanni et al., 1980). This presumably occurs by stimulating the metabolism of carcinogens through detoxification pathways to a greater extent than through toxification pathways (Conney, 1982).

Consistent with this longstanding hypothesis, Gelhaus et al. (2011) found that activation of AhR by pretreatment of H358 bronchial alveolar cells with TCDD caused a decrease in benzo[*a*]pyrene (BaP) adducts and an increase in BaP-glutathione adducts, the formation of which are catalyzed by GSTs (Kushman et al., 2007; Tang et al., 2010; Weng et al., 2005). Gelhaus et al. (2011) proposed that a detoxification pathway had been up-regulated as opposed to an activation pathway that had been down-regulated. The AhR thus appears to be involved in both toxification and detoxification events of several exogenous carcinogens. While there may be other pathways involved in lung carcinogenesis (Chiba et al., 2011), metabolic activation of smoke-borne carcinogens is almost universally accepted as a causative, initiating event.

A fully functional metabolic response to PAH exposure in airway epithelial cells involves the AhR-mediated upregulation of CYP1 enzymes and epoxide hydrolase, metabolism of PAHs to phenols, dihydrodiols and tetraols, phase II conjugation catalyzed by sulfotransferases and glucuronosyl transferases (Bernier et al., 1996; Nishimura and Naito, 2006; Vietri et al., 2003; Weibel et al., 1986), and phase III transport of the metabolites out of the cell mediated by transporters such as AhR-regulated ABCG2 (Ebert et al., 2005; 2007; Leslie et al., 2005; Tan et al., 2010). It is the intermediates in the pathway, the dihydrodiol epoxides, and possibly other reactive metabolites (Xue and Warshawsky, 2005), that give rise to the mutagenic DNA adducts. Thus, while a complete loss of AhR expression and function would prevent PAH-induced lung carcinogenesis, a reduced, but not totally abolished, level of AhR could actually lead to incomplete metabolism and transport of PAH carcinogens, reduced levels of GSTs that inactivate mutagenic dihydrodiol epoxides, and

consequently increased levels of the mutagenic intermediates that initiate cancers of the lung and other tissues.

Given the potentially pivotal role of the AhR in the development of lung and other cancers, there have been a number of studies focusing on genetic polymorphisms of the AhR and their potential roles in cancer susceptibility (Cauchi et al., 2001; Chen et al., 2009; Kawajiri et al., 1995; Kim et al., 2007; Li et al., 2013; Ng et al., 2010; Shin et al., 2008). Based on our previous studies (Englert et al., 2012), we contend that one of the most important genetic polymorphisms of the human *AHR* gene is the (GGGGC)_n repeat (with *n* = 2, 4, 5 or 6) in the proximal promoter region, which has essentially been overlooked for technical reasons. The TATA-less promoter of the human *AHR* gene is dependent on specificity protein (Sp) transcription factors for promoter activation. We hypothesize that this (GGGGC)_n repeat polymorphism, which is in the Sp transcription factor binding region of the *AHR* gene proximal promoter, influences inter-individual differences in susceptibility to PAH-induced carcinogens via its effect on AhR expression. As an initial step in determining the potential role of the (GGGGC)_n repeat polymorphism in human cancers, this study was undertaken to investigate the incidence of alleles containing varying numbers of the (GGGGC)_n repeats in DNA from patients with cancers of the lung, colon, or breast.

2. Materials and Methods

2.1. Cell culture and media

The NCI-H292 lung cancer cell line was obtained from the American Type Culture Collection (ATCC; Manassas, VA), and these cells were cultured in RPMI1640 medium (with phenol red) supplemented with 100 μM non-essential amino acids, 2 mM L-glutamine, 100 U/mL penicillin, 100 μg/mL streptomycin, and 10% fetal bovine serum (FBS; Sigma, Saint Louis, MO). MCF-7 cells were those used in our previous studies (Spink et al., 2003; Spink et al 2012), and they were cultured in DF₅, which consists of DMEM (with phenol red) supplemented with 100 μM non-essential amino acids, 2 mM L-glutamine, 100 U/mL penicillin, 100 μg/mL streptomycin, 10 ng/mL human recombinant insulin (Gibco, Life Technologies, Grand Island, NY), and 5% FBS. The colon adenocarcinoma cell line, Caco-2 (from ATCC), was cultured in DF₁₀, which differed from DF₅ in that it contained 10% FBS and no added insulin. All cultures were maintained at 37 °C in humidified air containing 5% CO₂. Ethoxyresorufin-O-deethylase (EROD) assays for all cell lines were carried out in DC₁₀ medium, which differed from DF₁₀ in that it contained 10% Cosmic Calf Serum (Hyclone, Logan, UT) rather than FBS, and it did not contain phenol red.

2.2 RNA isolation and real-time PCR

For determination of mRNA levels in NCI-H292, Caco-2, and MCF-7 cells, confluent cultures in 6-well plates were treated with 10 nM TCDD or 0.1% DMSO vehicle for 48 h. Total RNA was then isolated and reverse-transcribed, and levels of CYP1A1, CYP1B1, and 36B4 mRNA were quantified by real-time PCR using the primers and conditions previously described (Spink et al., 2003).

2.3 Western immunoblots of microsomal CYP1A1 and CYP1B1

Confluent cultures of NCI-H292, Caco-2, and MCF-7 cells in 10-cm dishes were treated with 10 nM TCDD or 0.1% DMSO vehicle for 48 h. Microsomes were prepared as described (Spink et al., 1997) and were suspended in 10 mM Tris-HCl, pH 7.4, 150 mM KCl, 20% glycerol. Microsomal proteins (30 µg/lane) were resolved with 10% Bis-Tris gels (NuPage; Invitrogen, Life Technologies) and blotted onto Invitrolon PVDF membranes (Invitrogen). Blots were probed with anti-CYP1A1 (H-70; Santa Cruz Biotechnology, Dallas, TX) or CYP1B1 (H-105; Santa Cruz Biotechnology) antibodies as described (Spink et al., 2003) and detected using the SuperSignal West Pico Chemiluminescent Substrate (Pierce, Thermo Fisher, Rockford, IL) and the ChemiDoc imaging system (BioRad, Hercules, CA).

2.4. EROD assay

EROD assays were carried out with cultures in 96-well plates as previously described (Spink et al., 2009). In brief, cells in their respective media were treated with 10 nM TCDD (Cambridge Isotope Laboratories, Andover, MA) or 0.1% dimethyl sulfoxide (DMSO) vehicle for 48 h. Conversion of the CYP1 substrate, ethoxyresorufin (Sigma), to resorufin (Sigma) was monitored after 30 min by measuring fluorescence at 590 nm elicited by excitation at 535 nm. EROD activities were normalized to total protein content as measured using the BCA protein assay (Thermo Fisher).

2.5. Luciferase assays of the promoters of CYP1A1, CYP1B1, and polymorphic AHR

The luciferase constructs containing approximately 1450 bp of the human CYP1A1 promoter, pHu-1A1-FL, or 1146 bp of the human *CYP1B1* promoter, p1B1Fluc, were those used in our previous studies (Spink et al., 2003; 2008). For the determination of *in vitro* promoter activities of the constructs, cells were seeded in 48-well plates in their respective medias and were transfected when the cultures were at approximately 60% confluence with 50 ng/well of the phRG-TK *Renilla* control vector, 150 ng/well of pHu-1A1-FL or p1B1Fluc, and 0.66 µL/well ExGen500 (Fermentas, Thermo Fisher, Rockford, IL) according to the manufacturer's protocol. After 24 h, cells were treated with 10 nM TCDD or the DMSO vehicle (0.1%) and assayed for luciferase activity 48 h after treatment using the Dual-Luciferase Reporter 1000 Assay System as recommended (Promega). Luminescent signals were recorded using a Lumat LB9501 luminometer (Berthold, Germany), and firefly luciferase activities were normalized to *Renilla* luciferase activities (Spink et al., 2009).

A deletion construct of the *AHR* promoter-luciferase reporter plasmid, pGL3-hAhRP, containing 5640 bp of the AhR promoter, (GGGGC)₄ in the proximal promoter region, and 159 bp corresponding to the untranslated region of the AhR mRNA, was prepared by restriction endonuclease digestion with *KpnI* and *ApaI* (Fermentas) as described (Wolff et al., 2001). The resulting construct, referred to as AhR (-120)₄, in which the subscript refers to the number of GGGGC-repeats, encompasses 120 bp of the proximal *AHR* promoter region and 159 bp of the region corresponding to the mRNA. To prepare a construct containing (GGGGC)₂, two repeats, or 10 bp, were removed from the AhR (-120)₄ construct using the QuikChange II Site-Directed Mutagenesis Kit (Stratagene, Agilent Technologies, Santa Clara, CA) as described (Englert et al., 2012), and is referred to as

AhR (-120)₂. This strategy was also used to prepare a construct containing the single nucleotide polymorphism (SNP) C/A: (GGGGC)₂GGGGAGGGGC, in the AhR (-120)₄ construct, which is referred to as AhR (-120)_{C/A}. The oligonucleotides used for site-directed mutagenesis are summarized in Supplementary Table 1. After verification of the inserts by nucleotide sequencing, they were subcloned into the original vector using *NotI*/*BglII* (Fermentas) to ensure identical luciferase and surrounding sequences. To obtain constructs containing (GGGGC)₅ and (GGGGC)₆, genomic DNA (gDNA) from a neonate heterozygous for (GGGGC)_{5/6} was amplified using the primers, AhRseq-F/R (Supplementary Table 1), and subsequent digestion with *PauI*/*ApaI* (Fermentas) for cloning into pGL3-hAhRP. The AhR (-120)₅ and AhR (-120)₆ constructs containing (GGGGC)₅ and (GGGGC)₆, respectively, were obtained by deletion subcloning of the newly prepared, pGL3-hAhRP constructs containing (GGGGC)₅ and (GGGGC)₆, using *KpnI* and *ApaI* as described above. The DNA sequences of the inserts from all clones were verified by sequencing.

For the determination of *in vitro* promoter activities of the constructs, cells were seeded in 48-well plates in their respective media, and were transfected when the cultures were at approximately 60% confluence, with 50 ng/well of the phRG-TK *Renilla* control vector, 150 ng/well of AhR (-120)₂, AhR (-120)₄, AhR (-120)₅, AhR (-120)₆, or AhR (-120)_{C/A}, and 0.66 μ L/well ExGen500 (Fermentas). After 48 h, the luciferases were assayed as described above for determination of the CYP1A1 and CYP1B1 promoter activities.

2.6. Sources of human DNA and isolation of gDNA

Surplus DNA derived from bloodspots of New York State neonates was obtained from the Newborn Screening Program of New York State Department of Health and represents an expanded data set from our previous study (Englert et al., 2012). Frozen non-tumor lung tissue from lung cancer patients and formalin-fixed, paraffin-embedded (FFPE) tumor tissue from breast and colon cancer patients were obtained from the National Cancer Institute's Cooperative Human Tissue Network (CHTN). The identities of all tissue donors were anonymous. Studies with the neonatal bloodspots and the human tissue samples were approved by the Institutional Review Board of the New York State Department of Health. FFPE or frozen tissue from cancer patients and matching non-tumor tissue from individual patients with various types of cancer were also obtained from CHTN. A section of tissue adjacent to that used for DNA isolation was fixed and stained with hematoxylin and eosin. The slides were subjected to independent pathologic examinations to confirm tumor or non-tumor status. Isolation of DNA from frozen lung tissue was accomplished by use of the DNeasy tissue kit (Qiagen, Valencia, CA) or the PureGene DNA Purification System (Gentra, Minneapolis, MN). DNA from FFPE tissue was isolated using the REPLI-g FFPE kit (Qiagen), but without the ligation and whole-genome amplification steps, and with an extended overnight digestion at 60°C with proteinase K. This was followed by the addition of a second aliquot of proteinase K and digestion for an additional 24 h.

2.7. Determination of (GGGGC)_n polymorphic repeats in gDNA from neonates, and lung, colon, and breast cancer patients

Polymerase chain reaction (PCR) of gDNA was performed using recombinant *Taq* DNA polymerase buffered with (NH₄)₂SO₄ (Fermentas), gene-specific primers, 6% propylene glycol (Zhang et al., 2009), 1.75 to 2.5 mM MgCl₂, and TaqStart Antibody (Clontech Laboratories, Mountain View, CA). Thermal cycling was performed using the Gene Amp 9600 PCR System (Applied Biosystems, Foster City, CA), employing a combination of slowdown (Frey et al., 2008) and touchdown methodologies. After initial denaturation at 95 °C for 2 min, the touchdown segment included 10 cycles of denaturation at 95 °C for 30 s, annealing at 70 °C for 30 s with a drop of 1 °C/cycle, and extension at 72 °C for 20–70 s. The slowdown segment consisted of 41–45 cycles of denaturation at 95 °C for 30 s, a decrease in temperature from 95 °C to 57 °C at 1.9 °C/s, annealing at 57 °C for 30 s, and extension at 72 °C for 20–70 s, followed by a final extension of 5 min at 72 °C. Longer extension times were used for larger amplicons, and an extension time of 70 s was used for amplification of potential microsatellites. For sequence analysis of PCR products containing (GGGGC)_n repeats, the primers spanning –125 to +66 bp relative to the transcriptional start site (TSS) (based on the sequence containing (GGGGC)₂), AhRseq-F and AhRseq-R were used (Supplementary Table 1). Protein was removed from the samples using StrataClean Resin (Stratagene, Agilent Technologies) prior to nucleotide sequencing. Genotypes containing SNPs were amplified at least two times for confirmatory sequencing. All genotypes were verified with respect to the number of (GGGGC)_n repeats by performing a separate PCR amplification using the AhRgel-F and AhRgel-R primers that span –57 to +8 nt relative to the TSS and subsequent heteroduplex analysis (Barros et al., 2005; Gtari et al., 2007) of products separated by polyacrylamide gel electrophoresis (PAGE). Differing concentrations of DNA and MgCl₂ were used in this second amplification to ensure that both alleles were amplified. Product sizes using AhRgel-F/R primers for (GGGGC)₂, (GGGGC)₄, (GGGGC)₅, and (GGGGC)₆ were 65, 75, 80, and 85 (bp), respectively.

For heteroduplex analysis, 3-μL aliquots of the PCR products obtained above were applied to NuPAGE 20% Tris-borate polyacrylamide gels (Life Technologies) or 17.4% acrylamide (Sigma)/0.6% bisacrylamide (BioRad, Hercules, CA) (v/v) Tris-borate gels. Gels were stained with ethidium bromide, and amplified gDNA or constructs, which had previously been sequenced to determine the number of repeats, were used as allele-specific markers during PAGE (Miri et al., 2012). To prepare a template containing (GGGGC)₃ for use as a standard, initially, a deletion construct of pGL3-hAhRP containing 881 bp of the AhR promoter containing (GGGGC)₄, referred to as AhR (-881)₄, was prepared by using restriction endonuclease digestion with *Kpn*I and *Sau*I (Fermentas) as described (Wolff et al., 2001). To obtain a construct containing (GGGGC)₃, GGGGC, or 5 bp, was removed from the AhR (-881)₄ construct using site-directed mutagenesis as described in Section 2.5, and is referred to as AhR (-881)₃, using oligonucleotides as described in Supplementary Table 1.

2.8. Analysis of non-human primate gDNA

Samples of primate gDNA from *Macaca nemestrina*, *Macaca fascicularis*, *Macaca mulatta*, *Pan troglodytes*, and *Gorilla gorilla* were obtained from several sources (see

Acknowledgments). The DNA was amplified by PCR for sequencing of regions of the DNA homologous to the human repeat polymorphism using the same conditions that are described in Section 2.7 using the primers, AhRseq-F and AhRseqP-R (Supplementary Table 1).

2.9. Statistical analysis

Data from in vitro experiments with cultured cells were subjected to ANOVA followed by the Bonferroni's multiple comparison test (Sigma Stat; Systat Software, San Jose, CA). Allelic frequency data were analyzed using omnibus χ^2 -tests or Fisher's exact test across diagnostic groups by sequence repeat, followed by pairwise χ^2 -tests or Fisher's exact tests between diagnostic groups, as appropriate. A Bonferroni correction was applied to accommodate type-1 error inflation associated with multiple pairwise comparisons. Statistical significance was defined as $P < 0.007$ for pairwise two-tailed comparisons within group (i.e., 0.05/7 for each sequence repeat). The SAS v. 9.3 statistical software package was used for analysis (SAS Institute, Inc., Cary, NC).

3. Results

3.1. Induction of CYP1A1 and CYP1B1 in human cancer cells

Ah-responsiveness for CYP1 induction was determined in representative lung, colon, and breast cancer cell lines, NCI-H292, Caco-2, and MCF-7, respectively, by measuring CYP1A1 and CYP1B1 mRNAs and proteins, EROD activity, and *CYP1A1* and *CYP1B1* promoter activities and after cultures were exposed to 10 nM TCDD or the DMSO vehicle. Each of the three cell lines showed a high level of induction of CYP1A1 mRNA after exposure to TCDD relative to control (Fig. 1A), with mean absolute CYP1A1 mRNA levels of 13.1, 34.3 and 13.5 amol/ μ g RNA in NCI-H292, Caco-2 and MCF-7 cells after exposure to TCDD. Western blots of microsomal proteins showed induction of CYP1A1 protein in each of the three cell lines. All three cell lines also showed high level of induction of CYP1B1 mRNA after exposure to TCDD relative to control (Fig. 1B); however, the mean absolute CYP1B1 mRNA levels after TCDD exposure were 3.91, 0.764 and 34.2 amol/ μ g RNA in NCI-H292, Caco-2 and MCF-7 cells, respectively. The low absolute levels of CYP1B1 mRNA in NCI-H292 and Caco-2 cells after exposure to TCDD appear to be consistent with the fact that induction of CYP1B1 protein was not observed by Western blot in either cell line (Fig. 1C). Despite 8-fold and 30-fold induction, respectively, CYP1B1 mRNA levels in NCI-H292 and Caco-2 cells did not appear to reach the threshold that is required for translation into detectable amounts of CYP1B1 protein, whereas CYP1B1 protein was readily detected in MCF-7 cells (Fig. 1C).

Each of the three cell lines expressed TCDD-inducible EROD activity (Fig. 1D), which is consistent with the observed induction of CYP1A1 mRNA and protein. The EROD assay measures predominately CYP1A1 activity, although CYP1B1 may also contribute to some extent. AhR-mediated transcriptional activation of *CYP1A1* and *CYP1B1* constructs was investigated in cells transfected with luciferase constructs containing *CYP1A1* and *CYP1B1* promoters, which were subsequently exposed to TCDD or vehicle. The TCDD-induced *CYP1A1* promoter activity was significantly increased in NCI-H292 and MCF-7 cells in

comparison with that of the vehicle control cultures (Fig. 1E). Despite the fact that the endogenous CYP1B1 gene appears to be largely silenced in NCI-H292 and Caco-2 cells, TCDD-induced CYP1B1 promoter activity was significantly increased in all three cell lines (Fig. 1F). The results for CYP1 mRNAs, proteins, EROD activity, and *CYP1* promoter activities, when taken together (Fig. 1), indicate that each of the three cell lines has a highly functional Ah response, and each of the cell lines would therefore be appropriate for studies of AhR expression.

3.2. Effects of the (GGGGC)_n repeats and polymorphic sequences on AhR promoter activity

Studies of AhR promoter activity were initiated using the AhR proximal promoter luciferase constructs that contained 2, 4, 5, or 6 GGGGC-repeats and the C/A SNP, (GGGGC)₂GGGGCAGGGGC (Fig. 2A). Since AhR promoter activity was measured, rather than CYP1 promoter activity, TCDD was not included in these experiments. After transfection of these constructs in the three cell lines, the luciferase activity for the construct containing (GGGGC)₂ was less than half of that observed for the parental (control) construct containing (GGGGC)₄ (Fig. 2B, 2C, and 2D). In MCF-7 and Caco-2 cells, the luciferase activity of the construct containing (GGGGC)₅ was increased relative to that of the construct containing (GGGGC)₄. The luciferase activity of the construct containing (GGGGC)₆ was not significantly different from that of the construct containing (GGGGC)₄ in the three cell lines, and only in MCF-7 cells was the activity of the construct containing (GGGGC)_{C/A} relative to (GGGGC)₄ significantly reduced.

3.3. Heteroduplex analysis of the polymorphic (GGGGC)_n repeats

Previous attempts in our laboratory to amplify regions of DNA spanning the polymorphic (GGGGC)_n repeat were unsuccessful due to the very high GC-content of the sequence. We therefore investigated a variety of PCR conditions, primers, and additives, and we found that the DNA was correctly amplified when the denaturant, propylene glycol (Zhang et al., 2009), was included in the PCR mix. Portions of the AhR promoter were amplified from surplus gDNA from New York State neonates, and the products were subsequently sequenced. In order to confirm the number of repeats, a smaller PCR product containing the (GGGGC)_n repeat was amplified and resolved by PAGE for heteroduplex analysis. Unambiguous identification of the polymorphic forms was not possible using separation with agarose gels (Fig. 3A); however, separation and identification were readily accomplished when the PCR products were resolved on 18–20% PAGE gels (Fig. 3B). The rare (GGGGC)₂ and (GGGGC)₆ repeats were only found in gDNA from heterozygous individuals. Heterozygosity was easily identified when the amplified DNA was analyzed on PAGE gels due to the presence of slower-migrating heteroduplexes, which were not present in the amplified DNA from (GGGGC)₄ or (GGGGC)₅ homozygous individuals. When PCR products from (GGGGC)₄ and (GGGGC)₅ homozygous individuals were mixed in equal amounts, melted at 95 °C and cooled prior to analysis, the migration pattern of the DNA was identical to that of PCR product from the heterozygous (GGGGC)_{4/5} individual (Fig 3C, lanes denoted as “4+5” and “4/5”, respectively). When the two heteroduplex bands from the heterozygous (GGGGC)_{4/5} individual were excised from the gel and reamplified, all bands, including those representing homoduplexes, were reproduced.

These results indicate that these slower-migrating bands do not represent expanded tandem repeats; rather, they are consistent with DNA structures that are formed when homologous PCR products that differ in length, due to unequal numbers of GGGGC repeats, hybridize when DNA from heterozygous individuals is amplified. The presence of these easily distinguished, slower-migrating forms is commonly exploited for heteroduplex analysis of microsatellite variability (Barros et al., 2005; Gtari et al., 2007). We then performed heteroduplex analysis by PCR amplification of the luciferase constructs containing 2, 4, 5, and 6 GGGGC-repeats (Section 2.5) as templates, including a construct that was engineered to contain a (GGGGC)₃ repeat (Section 2.7). The homoduplex products from these amplifications were gel-purified and mixed using all possible pair-wise combinations in approximately equal picomolar amounts, heat-denatured, and hybridized by quickly cooling. Fig. 3D shows the migration patterns of heteroduplex PCR products formed from these combinations.

3.4. Analysis of the polymorphic (GGGGC)_n repeat in tumor and non-tumor tissue

To determine whether the polymorphic (GGGGC)_n repeats are a result of microsatellite changes, gDNA from tumor and matching non-tumor tissue from 16 individuals was amplified and analyzed by heteroduplex analysis (Fig. 4). The PCR product of DNA from microsatellite expansions would be expected to form heteroduplex bands that differ in migration pattern from those of inherited alleles (Barros et al., 2005). The results of independent pathologic examinations for tumor or non-tumor status of the tissues and pathology reports from CHTN are summarized in Supplementary Table 2. The molecular events involved in carcinogenesis did not result in changes in the number of repeats, nor in changes of migration of any of the heteroduplex forms in these specimens, indicating that the observed (GGGGC)_n repeats are inherited alleles and not the results of microsatellite deletions or replications. In addition, the DNA from 29 tumor tissues from individuals with various types of cancer were analyzed for microsatellite changes, and found to contain only 2, 4, and 5 GGGGC-repeats (Fig. 5). These tissues were also verified to be tumor tissue by independent pathologic examinations, and the results and CHTN pathology reports are summarized in Supplementary Table 3.

3.5. Analysis of allele frequencies of the polymorphic (GGGGC)_n repeat in lung, colon, and breast cancer patients

The allelic frequencies of the polymorphic (GGGGC)_n repeats in blood spots from New York State neonates and tissues from cancer patients were determined by sequencing of PCR products and were confirmed by separate amplifications with different primers and subsequent heteroduplex analysis. The genotypes determined using the two methods were in agreement with respect to the number of GGGGC repeats, with the one exception of a neonate with the very unusual genotype of (GGGGC)_{4/6}, and harboring a rare G/T SNP (Table 1). For this individual, we could not amplify the (GGGGC)₆ allele using the primers for heteroduplex analysis. However, this individual also had a second, rare SNP in the (GGGGC)₆ allele homologous to the terminal end of the reverse primer used for heteroduplex analysis, which most likely prevented the efficient amplification of the (GGGGC)₆ allele. Table 1 is a summary of all individuals we found that harbor SNPs in the polymorphic region. The remaining SNPs that were identified were from 6 out of a total of

327 individuals who harbored the C/A SNP, (GGGGC)₂GGGG(C/A)GGGGC, which was present only in African Americans and Hispanics, including one individual who was homozygous for the C→A SNP. The demographics for neonates and lung, colon, and breast cancer patients are summarized in Table 2. The diagnoses for the cancer groups are summarized in Table 3. Allelic frequencies of the polymorphic repeat for all populations are presented in Fig. 6. The distribution of *P*-values and numbers of alleles for omnibus and pair-wise bivariate tests of sequence repeats and diagnostic group are presented in Supplementary Table 4. The results indicate that, in all populations examined, the most prominent allele was (GGGGC)₄, followed in incidence by (GGGGC)₅. In none of the 177 cancer patients was the (GGGGC)₆ allele observed; however, (GGGGC)₆ was also quite low in incidence in New York State neonates, with only three alleles observed in 150 newborns. The allele frequencies of (GGGGC)₂ in lung, colon, and breast cancer patients and neonates were 5.32%, 2.27%, 1.28% and 0.67%, respectively. The allele frequency of (GGGGC)₂ in lung cancer patients was 8-fold greater than that found in neonates from New York State (*P*<0.002).

3.6. Analysis of DNA from non-human primates in the region homologous to the polymorphic (GGGGC)_n repeat in humans

DNA from five species of primates was sequenced in the region homologous to that spanning the polymorphic (GGGGC)_n repeat in humans in order to determine whether similar polymorphisms occur in closely related species. Eight individuals of *Macaca mulatta* had identical sequences containing (GGGGC)₂, which agreed with the sequence in the Ensembl database (Flicek et al, 2014) for this species. Two other species of macaque showed the same sequence in the region of (GGGGC)₂. The DNA from 8 individuals of the common chimpanzee contained (GGGGC)₂ following an additional 5-base sequence, GGGGA, which also agreed with the sequence for the chimpanzee in Ensembl. The sequence of this region in western gorilla was indeterminate in Ensembl; therefore, we sequenced the DNA from 6 individuals of gorilla and found (GGGGC)_n repeats with *n* = 2,4,5,6,7 and 8. These results are summarized in Table 4, and the alignment of these sequences can be found in Supplementary Fig. 1.

4. Discussion

In several previous studies of polymorphisms of the *AHR* gene, the focus was on coding sequence changes that result in amino acid changes in the receptor protein, which may result in functional changes or alterations in ligand binding (Cauchi et al., 2001; Rowlands et al., 2010; Tiido et al., 2007; Kawajiri et al., 1995; Ng et al., 2010; Shin et al., 2008). There have also been a number of studies of polymorphisms of the *AHR* gene in which the sequences flanking the GGGGC-repeat were included (Cauchi et al., 2001; Racky et al., 2004; Ng et al., 2010; Li et al., 2013; Fukushima-Uesaka et al., 2004); however, in only two of these studies was the (GGGGC)_n repeat polymorphism reported (Li et al., 2013; Fukushima-Uesaka et al., 2004). In an analysis of polymorphisms of the *AHR* gene in a Japanese population, the GGGGC short tandem repeat polymorphism was identified, and the allele frequencies of (GGGGC)_n were reported as *n* = 2>4,5,6 (Fukushima-Uesaka et al., 2004). In another study of the Japanese population, allele frequencies of (GGGGC)_n were specified as

$n = 3 > 2 \gg 1, 4, 5$ (Li et al., 2013). The results of these two studies are not consistent with each other or with our findings. Our data indicate that the most prominent allele is (GGGGC)₄, followed in incidence by (GGGGC)₅; (GGGGC)₂ and (GGGGC)₆ were found to be rare alleles, and the (GGGGC)₁ and (GGGGC)₃ alleles were not observed. In our previous study of New York State neonates (Englert et al., 2012), we found that the individuals of Asian descent also showed (GGGGC)₄ as the most prominent allele, followed in incidence by (GGGGC)₅. The reasons for the discrepancies among our results and those of Fukushima-Uesaka et al. (2004) and Li et al. (2013) are not apparent.

In our study, we observed no evidence of microsatellite instability of the (GGGGC)_n repeat. The results reported here also indicate that the alleles of the (GGGGC)_n polymorphism are inherited through the germline. This is in contrast with the possibility that they result from microsatellite additions or deletions often observed in cancer, since the polymorphism was observed in non-tumor tissues from lung and in bloodspots from neonates, and DNA from matching tumor and non-tumor tissues did not show differences in the number of repeats. An advantage of using PAGE for the analysis of microsatellites is that minor microsatellite forms present in cancer tissue would be visible as heteroduplex bands that migrate differently from those of allelic heteroduplexes. Amplifications of DNA from a large set of tumor tissues did not show heteroduplex bands other than those arising from the alleles, (GGGGC)₂, (GGGGC)₄, or (GGGGC)₅. A disadvantage of the analysis of heteroduplexes by PAGE is that the C/A SNP present in the (GGGGC)₄-allele cannot be identified without nucleotide sequencing. The heteroduplex forms observed upon analysis by PAGE are the result of the hybridization of homologous PCR products that differ in one or more GGGGC-repeats when DNA from heterozygous individuals is amplified, melted, and rehybridized. The loop formed from the larger amplicon impedes migration during PAGE (Barros et al., 2005). Two bands are typically seen, which correspond to the looping of either the template or coding DNA strands.

A comparison study of human *AHR* proximal promoter sequences with those of other primate species was conducted to determine whether (GGGGC)_n repeats are unique to humans or are commonly observed among primates. Our results indicated that repeats of (GGGGC)_n in which $n > 2$ were readily observed in DNA from gorilla and humans, but not in numerous samples from four other primate species investigated in this study. These results are summarized in the context of the evolutionary relationships of primates in Supplementary Figure 2.

The question arises as to what potential impact these polymorphisms of the human *AHR* gene have on AhR expression and cancer susceptibility. To investigate this question we analyzed a series of NYS neonates and subjects with several human cancers. The most statistically significant association that was observed in cancer patients in comparison with the neonates was an increased incidence of (GGGGC)₂ in lung cancer patients. This finding is not the result of microsatellite changes, since the tissue we used to determine allele frequencies was from non-tumor lung tissue. The allelic frequency of (GGGGC)₂ was 8-fold greater in lung cancer patients in comparison with that of neonates, indicating that 10.6% of lung cancer patients in our study were heterozygous for this polymorphism. If our samples represent a random sampling of lung cancer patients, then these data suggest that as many as

24,000 of the 228,190 lung cancer patients who are diagnosed yearly in the United States (Siegel et al, 2013) have the rare, (GGGGC)₂ polymorphic form. While suggestive of a potential association of the (GGGGC)₂ form with lung cancer susceptibility, these results are limited, and suggest that a larger, case-controlled study should be undertaken with lung cancer patients and appropriately matched, cancer-free control patients.

We also found that the activity of the *AHR* luciferase construct containing the (GGGGC)₄ allele was increased in lung cancer cells in comparison with that of the construct containing the (GGGGC)₂ allele. Although transcription of heterologous luciferase constructs that are composed of “naked” DNA differs from that of gDNA associated with proteins and with possible epigenetic changes, these results suggest that *in vivo* transcription of the *AHR* allele with only 2 polymorphic repeats may be reduced. Indeed, we found in our previous study (Englert et al., 2012) that oligonucleotides containing the polymorphic repeats formed G-quadruplexes as measured by circular dichroism, and those with higher numbers of repeats displayed increased peak intensities. These spectra also suggested that the G-quadruplex formed with oligonucleotides containing the 2-repeat was of a qualitatively different conformation from the G-quadruplexes formed with oligonucleotides containing higher numbers of repeats.

The proximal promoter region of the human *AHR* gene is very GC-rich and is replete with putative Sp1 sites (Narayan et al., 1997; Matys et al, 2006), and overlapping with them are sequences that could potentially form G-quadruplexes. It appears that this arrangement within the *AHR* promoter is also observed in the promoters of many other genes (Todd and Neidle, 2008; Kumar et al., 2011), including that of the human *VEGF* gene (Sun et al., 2008). Using Sp1 ChIP-on-chip, Raiber et al. (2012) found that 77% of Sp1 binding sites are also putative G-quadruplex-forming sequences. Potential mechanisms involving G-quadruplexes that may affect transcription include the recruitment of activators or inhibitors of RNA polymerase. In addition, the formation of G-quadruplexes on the template strand may prevent transcription, whereas formation of G-quadruplexes on the coding strand may enhance transcription by keeping the DNA in a single-stranded, open conformation (Bochman et al., 2012). Sp1, which is known to bind to duplex DNA, was recently shown to bind to a single strand of DNA containing a G-quadruplex (Raiber et al., 2012). What is interesting about the human *AHR* gene is that it contains a polymorphism in this critical region of the proximal promoter. These polymorphic DNA forms would be expected to differ in Sp1 binding and the propensity to form G-quadruplexes.

The transcriptional activation of the *AHR* gene may differ depending on which of the polymorphic (GGGGC)_n forms is present through the processes discussed above. We hypothesize that reduced expression of AhR from the allele containing (GGGGC)₂ in lung may result in a decreased induction of CYP1A1 and CYP1B1, and consequently decreased rates of metabolism of carcinogens, particularly in smokers. Lung epithelial cells of smokers are exposed to much higher concentrations of carcinogenic PAHs than are cells of other tissues (Boffetta et al, 1997). We (Englert et al., 2011;2012; Spink et al., 2003; 2012) and others (Do et al., 2014; Oyama et al., 2007) have found that Ah-responsiveness for induction of CYP1 enzymes is highly correlated with the level of expression of AhR mRNA. Altered expression of other AhR-regulated genes, such as phase II GSTs and phase III transporters,

may also be associated with variations in AhR expression. Based on the effects of (GGGGC)_n polymorphic sequences on *AHR* promoter activity, we hypothesize that the effect of the (GGGGC)_n polymorphism is on the level of AhR expression, ultimately affecting a delicate balance between and metabolic activation and detoxification of carcinogens. However, since the (GGGGC)_n polymorphisms are heritable, different alleles may be genetically linked to other significant genetic elements that affect cancer incidence. The underlying mechanisms can only be determined through additional research, which should include studies of allelic expression in the context of biological function and cancer susceptibility.

We might posit that in tissues without direct exposure to high levels of carcinogens, such as the breast, the consequences of this polymorphism on AhR expression may be more subtle, or may affect one or more of the other roles of the AhR, such as its involvement in control of the cell cycle, immune function, and development (Abel and Haarmann-Stemman, 2010; Denison et al, 2011; Gomez-Duran et al., 2009; Head and Lawrence, 2009; Kerkvliet, 2009; Lahoti et al, 2014; Puga et al., 2009). As has been reported previously, the factors that impact AhR expression are multiple and complex, and are usually found to be cell-specific (Harper et al, 2006). Based on the studies reported here, we suggest that the *AHR* proximal promoter GGGGC-repeat polymorphism may have consequences for AhR expression and lung carcinogenesis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by National Institutes of Health grants CA081243 and CA170960 (to DCS), grant CA092596 (to XD), and a grant from the New York State Attorney General's office, grant N1407157 (to ES). The authors gratefully acknowledge use of the Wadsworth Center's Media and Tissue Culture Facility, the Biochemistry, Histopathology, and Applied Genomic Technologies Core Facilities, and the Molecular Diagnostics Laboratory. We thank Richard Cole, Director of the Wadsworth Center's Advanced Light Microscopy Core Facility, for helpful advice on fixation of frozen tissue. The *AHR* promoter plasmid, pGL3-hAhRP, was a generous gift from Drs. Sandra Wolff and Josef Abel of the IUF-Leibniz Research Institute for Environmental Medicine, Heinrich-Heine-University, Düsseldorf, Germany. DNA from 3 species of macaque was a generous gift of Dr. Roger Wiseman of the Genetics Services Unit of the Wisconsin National Primate Research Center, University of Wisconsin, Madison, WI. DNA from gorilla was a generous gift of Dr. Michael I. Jensen-Seaman, Department of Biological Sciences, Duquesne University, Pittsburgh, PA. DNA from chimpanzee was provided by Dr. Jerilyn Pecotte of the Southwest National Primate Research Center, San Antonio, TX (Biomaterials Distribution Program), which was funded by the National Center for Research Resources (P51 RR013986) and is currently supported by the Office of Research Infrastructure Programs/OD P51 OD011133.

Abbreviations

AhR	aryl hydrocarbon receptor
ATCC	American Type Culture Collection
BaP	benzo[<i>a</i>]pyrene
CHTN	Cooperative Human Tissue Network
CYP1	cytochrome P450 family 1

DMSO	dimethyl sulfoxide
E₂	17 β -estradiol
EROD	ethoxyresorufin- <i>O</i> -deethylase
FBS	fetal bovine serum
FFPE	formalin-fixed, paraffin-embedded
gDNA	genomic DNA
PAGE	polyacrylamide gel electrophoresis
PAH	polycyclic aromatic hydrocarbon
PCR	polymerase chain reaction
SNP	single nucleotide polymorphism
Sp	specificity protein
TCDD	2,3,7,8-tetrachlorodibenzo- <i>p</i> -dioxin
TSS	transcriptional start site

References

- Abel J, Haarmann-Stemmann T. An introduction to the molecular basics of aryl hydrocarbon receptor biology. *Biol Chem*. 2010; 391:1235–1248. [PubMed: 20868221]
- Barros P, Blanco MG, Boan F, Gomez-Marquez J. Heteroduplex analysis of minisatellite variability. *Electrophoresis*. 2005; 26:4304–4309. [PubMed: 16240297]
- Bernier F, Soucy P, Luu-The V. Human phenol sulfotransferase gene contains two alternative promoters: Structure and expression of the gene. *DNA Cell Biol*. 1996; 15:367–375. [PubMed: 8924211]
- Bochman ML, Paeschke K, Zakian VA. DNA secondary structures: stability and function of G-quadruplex structures. *Nat Rev Genet*. 2012; 13:770–780. [PubMed: 23032257]
- Boffetta P, Jourenkova N, Gustavsson P. Cancer risk from occupational and environmental exposure to polycyclic aromatic hydrocarbons. *Cancer Causes Control*. 1997; 8:444–472. [PubMed: 9498904]
- Cauchi S, Stucker I, Solas C, Laurent-Puig P, Cenee S, Hemon D, Jacquet M, Kremers P, Beaune P, Massaad-Massade L. Polymorphisms of human aryl hydrocarbon receptor (AhR) gene in a French population: relationship with CYP1A1 inducibility and lung cancer. *Carcinogenesis*. 2001; 22:1819–1824. [PubMed: 11698344]
- Chen D, Tian T, Wang H, Liu H, Hu Z, Wang Y, Liu Y, Ma H, Fan W, Miao R, Sun W, Qian J, Jin L, Wei Q, Shen H, Huang W, Lu D. Association of human aryl hydrocarbon receptor gene polymorphisms with risk of lung cancer among cigarette smokers in a Chinese population. *Pharmacogenet Genomics*. 2009; 19:25–34. [PubMed: 18818557]
- Chiba T, Uchi H, Yasukawa F, Furue M. Role of the arylhydrocarbon receptor in lung disease. *Int Arch Allergy Immunol*. 2011; 155(Suppl 1):129–134. [PubMed: 21646808]
- Conney AH. Induction of microsomal enzymes by foreign chemicals and carcinogenesis by polycyclic aromatic hydrocarbons: G. H. A. Clowes Memorial Lecture. *Cancer Res*. 1982; 42:4875–4917. [PubMed: 6814745]
- Conney AH. Enzyme induction and dietary chemicals as approaches to cancer chemoprevention: the Seventh DeWitt S. Goodman Lecture. *Cancer Res*. 2003; 63:7005–7031. [PubMed: 14612489]
- Denison MS, Soshilov AA, He G, DeGroot DE, Zhao B. Exactly the same but different: promiscuity and diversity in the molecular mechanisms of action of the aryl hydrocarbon (dioxin) receptor. *Toxicol Sci*. 2011; 124:1–22. [PubMed: 21908767]

- DiGiovanni J, Berry DL, Gleason GL, Kishore GS, Slaga TJ. Time-dependent inhibition by 2,3,7,8-tetrachlorodibenzo-p-dioxin of skin tumorigenesis with polycyclic hydrocarbons. *Cancer Res.* 1980; 40:1580–1587. [PubMed: 6768449]
- Do MT, Kim HG, Tran TT, Khanal T, Choi JH, Chung YC, Jeong TC, Jeong HG. Metformin suppresses CYP1A1 and CYP1B1 expression in breast cancer cells by down-regulating aryl hydrocarbon receptor expression. *Toxicol Appl Pharmacol.* 2014; 280:138–148. [PubMed: 25110054]
- Ebert B, Seidel A, Lampen A. Identification of BCRP as transporter of benzo[a]pyrene conjugates metabolically formed in Caco-2 cells and its induction by Ah-receptor agonists. *Carcinogenesis.* 2005; 26:1754–1763. [PubMed: 15917307]
- Ebert B, Seidel A, Lampen A. Phytochemicals induce breast cancer resistance protein in Caco-2 cells and enhance the transport of benzo[a]pyrene-3-sulfate. *Toxicol Sci.* 2007; 96:227–236. [PubMed: 17077187]
- Ema M, Ohe N, Suzuki M, Mimura J, Sogawa K, Ikawa S, Fujii-Kuriyama Y. Dioxin binding activities of polymorphic forms of mouse and human arylhydrocarbon receptors. *J Biol Chem.* 1994; 269:27337–27343. [PubMed: 7961644]
- Englert NA, Spink BC, Spink DC. Persistent and non-persistent changes in gene expression result from long-term estrogen exposure of MCF-7 breast cancer cells. *J Steroid Biochem Mol Biol.* 2011; 123:140–150. [PubMed: 21185374]
- Englert NA, Turesky RJ, Han W, Bessette EE, Spivack SD, Caggana M, Spink DC, Spink BC. Genetic and epigenetic regulation of AHR gene expression in MCF-7 breast cancer cells: role of the proximal promoter GC-rich region. *Biochem Pharmacol.* 2012; 84:722–735. [PubMed: 22728919]
- Flicek P, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S, Gil L, Giron CG, Gordon L, Hourlier T, Hunt S, Johnson N, Juettemann T, Kahari AK, Keenan S, Kulesha E, Martin FJ, Maurel T, McLaren WM, Murphy DN, Nag R, Overduin B, Pignatelli M, Pritchard B, Pritchard E, Riat HS, Ruffier M, Sheppard D, Taylor K, Thormann A, Trevanion SJ, Vullo A, Wilder SP, Wilson M, Zadissa A, Aken BL, Birney E, Cunningham F, Harrow J, Herrero J, Hubbard TJ, Kinsella R, Muffato M, Parker A, Spudich G, Yates A, Zerbino DR, Searle SM. Ensembl 2014. *Nucleic Acids Res.* 2014; 42:D749–755. [PubMed: 24316576]
- Frey UH, Bachmann HS, Peters J, Siffert W. PCR-amplification of GC-rich regions: 'slowdown PCR'. *Nat Protoc.* 2008; 3:1312–1317. [PubMed: 18714299]
- Fukushima-Uesaka H, Sai K, Maekawa K, Koyano S, Kaniwa N, Ozawa S, Kawamoto M, Kamatani N, Komamura K, Kamakura S, Kitakaze M, Tomoike H, Ueno K, Minami H, Ohtsu A, Shirao K, Yoshida T, Saijo N, Saito Y, Sawada J. Genetic variations of the AHR gene encoding aryl hydrocarbon receptor in a Japanese population. *Drug Metab Pharmacokinet.* 2004; 19:320–326. [PubMed: 15499202]
- Gelhaus SL, Harvey RG, Penning TM, Blair IA. Regulation of benzo[a]pyrene-mediated DNA- and glutathione-adduct formation by 2,3,7,8-tetrachlorodibenzo-p-dioxin in human lung cells. *Chem Res Toxicol.* 2011; 24:89–98. [PubMed: 21028851]
- Gomez-Duran A, Carvajal-Gonzalez JM, Mulero-Navarro S, Santiago-Josefat B, Puga A, Fernandez-Salguero PM. Fitting a xenobiotic receptor into cell homeostasis: how the dioxin receptor interacts with TGFbeta signaling. *Biochem Pharmacol.* 2009; 77:700–712. [PubMed: 18812170]
- Gtari M, Brusetti L, Cherif A, Boudabous A, Daffonchio D. Heteroduplex structures in 16S-23S rRNA intergenic transcribed spacer PCR products reveal ribosomal interoperonic polymorphisms within single *Frankia* strains. *J Appl Microbiol.* 2007; 103:1031–1040. [PubMed: 17897207]
- Harper PA, Riddick DS, Okey AB. Regulating the regulator: factors that control levels and activity of the aryl hydrocarbon receptor. *Biochem Pharmacol.* 2006; 72:267–279. [PubMed: 16488401]
- Head JL, Lawrence BP. The aryl hydrocarbon receptor is a modulator of anti-viral immunity. *Biochem Pharmacol.* 2009; 77:642–653. [PubMed: 19027719]
- Kawajiri K, Watanabe J, Eguchi H, Nakachi K, Kiyohara C, Hayashi S. Polymorphisms of human Ah receptor gene are not involved in lung cancer. *Pharmacogenetics.* 1995; 5:151–158. [PubMed: 7550366]
- Kerkvliet NI. AHR-mediated immunomodulation: the role of altered gene transcription. *Biochem Pharmacol.* 2009; 77:746–760. [PubMed: 19100241]

- Kim JH, Kim H, Lee KY, Kang JW, Lee KH, Park SY, Yoon HI, Jheon SH, Sung SW, Hong YC. Aryl hydrocarbon receptor gene polymorphisms affect lung cancer risk. *Lung Cancer*. 2007; 56:9–15. [PubMed: 17174437]
- Kumar P, Yadav VK, Baral A, Saha D, Chowdhury S. Zinc-finger transcription factors are associated with guanine quadruplex motifs in human, chimpanzee, mouse and rat promoters genome-wide. *Nucleic Acids Res*. 2011; 39:8005–8016. [PubMed: 21729868]
- Kushman ME, Kabler SL, Ahmad S, Doehmer J, Morrow CS, Townsend AJ. Protective efficacy of hGSTM1-1 against B[a]P and (+)- or (–)-B[a]P-7,8-dihydrodiol cytotoxicity, mutagenicity, and macromolecular adducts in V79 cells coexpressing hCYP1A1. *Toxicol Sci*. 2007; 99:51–57. [PubMed: 17525473]
- Lahoti TS, Hughes JM, Kusnadi A, John K, Zhu B, Murray IA, Gowda K, Peters JM, Amin SG, Perdew GH. Aryl hydrocarbon receptor antagonism attenuates growth factor expression, proliferation, and migration in fibroblast-like synoviocytes from patients with rheumatoid arthritis. *J Pharmacol Exp Ther*. 2014; 348:236–245. [PubMed: 24309559]
- Leslie EM, Deeley RG, Cole SP. Multidrug resistance proteins: role of P-glycoprotein, MRP1, MRP2, and BCRP (ABCG2) in tissue defense. *Toxicol Appl Pharmacol*. 2005; 204:216–237. [PubMed: 15845415]
- Li D, Takao T, Tsunematsu R, Morokuma S, Fukushima K, Kobayashi H, Saito T, Furue M, Wake N, Asanoma K. Inhibition of AHR transcription by NF1C is affected by a single-nucleotide polymorphism, and is involved in suppression of human uterine endometrial cancer. *Oncogene*. 2013; 32:4950–4959. [PubMed: 23208493]
- Maglott D, Ostell J, Pruitt KD, Tatusova T. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res*. 2005; 33:D54–58. [PubMed: 15608257]
- Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, Voss N, Stegmaier P, Lewicki-Potapov B, Saxel H, Kel AE, Wingender E. TRANSFAC and its module TRANSCOMP: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res*. 2006; 34:D108–110. [PubMed: 16381825]
- Miri M, Hemati S, Safari F, Tavassoli M. GGC(n) polymorphism of eRF3a/GSPT1 gene and breast cancer susceptibility. *Med Oncol*. 2012; 29:1581–1585. [PubMed: 22101789]
- Narayan VA, Kriwacki RW, Caradonna JP. Structures of zinc finger domains from transcription factor Sp1. Insights into sequence-specific protein-DNA recognition. *J Biol Chem*. 1997; 272:7801–7809. [PubMed: 9065444]
- Nebert DW, Dalton TP, Okey AB, Gonzalez FJ. Role of aryl hydrocarbon receptor-mediated induction of the CYP1 enzymes in environmental toxicity and cancer. *J Biol Chem*. 2004; 279:23847–23850. [PubMed: 15028720]
- Nebert DW, Gelboin HV. Substrate-inducible microsomal aryl hydroxylase in mammalian cell culture. II. Cellular responses during enzyme induction. *J Biol Chem*. 1968; 243:6250–6261. [PubMed: 5723465]
- Nebert DW, Roe AL, Dieter MZ, Solis WA, Yang Y, Dalton TP. Role of the aromatic hydrocarbon receptor and [Ah] gene battery in the oxidative stress response, cell cycle control, and apoptosis. *Biochem Pharmacol*. 2000; 59:65–85. [PubMed: 10605936]
- Ng CH, Janoo-Gilani R, Sipahimalani P, Gallagher RP, Gascoyne RD, Connors JM, Weber JP, Lai AS, Leach S, Le ND, Brooks-Wilson AR, Spinelli JJ. Interaction between organochlorines and the AHR gene, and risk of non-Hodgkin lymphoma. *Cancer Causes Control*. 2010; 21:11–22. [PubMed: 19821039]
- Nishimura M, Naito S. Tissue-specific mRNA expression profiles of human phase I metabolizing enzymes except for cytochrome P450 and phase II metabolizing enzymes. *Drug Metab Pharmacokinet*. 2006; 21:357–374. [PubMed: 17072089]
- Okey AB. An aryl hydrocarbon receptor odyssey to the shores of toxicology: the Deichmann Lecture, International Congress of Toxicology-XI. *Toxicol Sci*. 2007; 98:5–38. [PubMed: 17569696]
- Oyama T, Sugio K, Uramoto H, Kawamoto T, Kagawa N, Nadaf S, Carbone D, Yasumoto K. Cytochrome P450 expression (CYP) in non-small cell lung cancer. *Front Biosci*. 2007; 12:2299–2308. [PubMed: 17127240]

- Poland A, Glover E, Kende AS. Stereospecific, high affinity binding of 2,3,7,8-tetrachlorodibenzo-p-dioxin by hepatic cytosol. Evidence that the binding species is receptor for induction of aryl hydrocarbon hydroxylase. *J Biol Chem*. 1976; 251:4936–4946. [PubMed: 956169]
- Puga A, Ma C, Marlowe JL. The aryl hydrocarbon receptor cross-talks with multiple signal transduction pathways. *Biochem Pharmacol*. 2009; 77:713–722. [PubMed: 18817753]
- Racky J, Schmitz HJ, Kauffmann HM, Schrenk D. Single nucleotide polymorphism analysis and functional characterization of the human Ah receptor (AhR) gene promoter. *Arch Biochem Biophys*. 2004; 421:91–98. [PubMed: 14678789]
- Raiber EA, Kranaster R, Lam E, Nikan M, Balasubramanian S. A non-canonical DNA structure is a binding motif for the transcription factor SP1 in vitro. *Nucleic Acids Res*. 2012; 40:1499–1508. [PubMed: 22021377]
- Rogers J, Gibbs RA. Comparative primate genomics: emerging patterns of genome content and dynamics. *Nat Rev Genet*. 2014; 15:347–359. [PubMed: 24709753]
- Rowlands CJ, Staskal DF, Gollapudi B, Budinsky R. The human AHR: identification of single nucleotide polymorphisms from six ethnic populations. *Pharmacogenet Genomics*. 2010; 20:283–290. [PubMed: 20401977]
- Shimizu Y, Nakatsuru Y, Ichinose M, Takahashi Y, Kume H, Mimura J, Fujii-Kuriyama Y, Ishikawa T. Benzo[a]pyrene carcinogenicity is lost in mice lacking the aryl hydrocarbon receptor. *Proc Natl Acad Sci U S A*. 2000; 97:779–782. [PubMed: 10639156]
- Shin A, Shrubsole MJ, Rice JM, Cai Q, Doll MA, Long J, Smalley WE, Shyr Y, Sinha R, Ness RM, Hein DW, Zheng W. Meat intake, heterocyclic amine exposure, and metabolizing enzyme polymorphisms in relation to colorectal polyp risk. *Cancer Epidemiol Biomarkers Prev*. 2008; 17:320–329. [PubMed: 18268115]
- Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. *CA Cancer J Clin*. 2013; 63:11–30. [PubMed: 23335087]
- Spink BC, Bennett JA, Lostritto N, Cole JR, Spink DC. Expression of the aryl hydrocarbon receptor is not required for the proliferation, migration, invasion, or estrogen-dependent tumorigenesis of MCF-7 breast cancer cells. *Mol Carcinog*. 2012; 52:544–554. [PubMed: 22388733]
- Spink BC, Bennett JA, Pentecost BT, Lostritto N, Englert NA, Benn GK, Goodenough AK, Turesky RJ, Spink DC. Long-term estrogen exposure promotes carcinogen bioactivation, induces persistent changes in gene expression, and enhances the tumorigenicity of MCF-7 human breast cancer cells. *Toxicol Appl Pharmacol*. 2009; 240:355–366. [PubMed: 19619570]
- Spink DC, Katz BH, Hussain MM, Pentecost BT, Cao Z, Spink BC. Estrogen regulates Ah responsiveness in MCF-7 breast cancer cells. *Carcinogenesis*. 2003; 24:1941–1950. [PubMed: 12970067]
- Spink DC, Spink BC, Cao JQ, Gierthy JF, Hayes CL, Li Y, Sutter TR. Induction of cytochrome P450 1B1 and catechol estrogen metabolism in ACHN human renal adenocarcinoma cells. *J Steroid Biochem Mol Biol*. 1997; 62:223–232. [PubMed: 9393958]
- Spink DC, Wu SJ, Spink BC, Hussain MM, Vakharia DD, Pentecost BT, Kaminsky LS. Induction of CYP1A1 and CYP1B1 by benzo(k)fluoranthene and benzo(a)pyrene in T-47D human breast cancer cells: roles of PAH interactions and PAH metabolites. *Toxicol Appl Pharmacol*. 2008; 226:213–224. [PubMed: 17919675]
- Sun D, Liu WJ, Guo K, Rusche JJ, Ebbinghaus S, Gokhale V, Hurley LH. The proximal promoter region of the human vascular endothelial growth factor gene has a G-quadruplex structure that can be targeted by G-quadruplex-interactive agents. *Mol Cancer Ther*. 2008; 7:880–889. [PubMed: 18413801]
- Tan KP, Wang B, Yang M, Boutros PC, Macaulay J, Xu H, Chuang AI, Kosuge K, Yamamoto M, Takahashi S, Wu AM, Ross DD, Harper PA, Ito S. Aryl hydrocarbon receptor is a transcriptional activator of the human breast cancer resistance protein (BCRP/ABCG2). *Mol Pharmacol*. 2010; 78:175–185. [PubMed: 20460431]
- Tang SC, Sheu GT, Wong RH, Huang CY, Weng MW, Lee LW, Hsu CP, Ko JL. Expression of glutathione S-transferase M2 in stage I/II non-small cell lung cancer and alleviation of DNA damage exposure to benzo[a]pyrene. *Toxicol Lett*. 2010; 192:316–323. [PubMed: 19900515]

Highlights

- The *AHR* proximal promoter contains a repeat polymorphism, (GGGGC)_n, where $n=4>5\gg 2,6$
- Matched tumor and non-tumor DNA did not show (GGGGC)_n microsatellite instability
- *AHR* promoter activity of a construct with (GGGGC)₂ was lower than that of (GGGGC)₄
- The frequency of (GGGGC)₂ in lung cancer patients was 8-fold higher than in neonates
- The (GGGGC)₂ allele may be associated with lung cancer susceptibility

Figure 1.

Ah-responsiveness in NCI-H292, Caco-2, and MCF-7 cancer cells. Panels **A** and **B**: CYP1A1 (**A**) and CYP1B1 (**B**) mRNA levels were determined in the three cell lines exposed to 10 nM TCDD (gray bars) or DMSO control (0.1%; black bars) for 48 h; $n = 3$. Panel **C**: Western immunoblots of microsomes isolated from the three cell lines after cultures were exposed to 10 nM TCDD or vehicle for 48 h; blots were probed with anti-CYP1A1 or -1B1 antibodies as indicated. Standards (Std) are microsomes containing cDNA-expressed CYP1A1 and CYP1B1 (Gentest, BD Biosciences, San Jose, CA). Panel **D**: EROD activity was assayed in cancer cells after 48-h exposure to 10 nM TCDD (gray bars) or DMSO control (0.1%; black bars); $n = 5$. Panels **E** and **F**: Cancer cells were transfected with luciferase constructs containing the CYP1A1 (**E**) or CYP1B1 (**F**) promoters. Cells were then exposed to 10 nM TCDD (gray bars) or DMSO control (0.1%; black bars) for 48 h prior to luciferase assays; $n = 4$. All determinations are presented as fold relative to the control and are means $\pm$ standard error; **, $P < 0.01$; ***, $P < 0.001$.

Figure 2.

Effects of the number of (GGGGC)_n repeat sequences on *AHR* promoter activity. Panel **A**: Depiction of the constructs containing the repeat polymorphisms found in the *AHR* proximal promoter fused to the luciferase coding sequence (*luc*). Shown are the five polymorphic forms of the promoter containing (GGGGC)_n, where n = 2, 4, 5, 6, and the rare C/A SNP: gatctgggc(GGGGC)₂GGGGAGGGGCcggtgagggg. Consensus Sp binding sites are shown as vertical black lines. The transcriptional start site is indicated (TSS). Panels **B**, **C**, and **D**: Constructs containing the polymorphisms represented in Panel **A**, AhR (-120)₂, AhR (-120)₄, AhR (-120)₅, AhR (-120)₆, or AhR (-120)_{C/A}, indicated by 2, 4, 5, 6, and 4 C/A, respectively, were transfected into the cancer cell lines, as indicated; after 48 h, the luciferase assays were performed, and the data expressed relative to the control vector, AhR (-120)₄; n = 7 or 8. All activities are presented as fold relative to the control and are means ± standard error; *, *P* < 0.05; ***, *P* < 0.001.

Figure 3.

AHR gene promoter polymorphic repeats and electrophoretic separation of PCR products. Homozygous (GGGGC)₄ or ₅ (**4**, **5**) or heterozygous (GGGGC)_n-repeat standards (**2/4**, **5/6**, **4/5**) were separated on a 2.5% agarose gel (**A**) or 20% polyacrylamide gel (**B**). **C**) Homozygous (GGGGC)₄ or ₅ (**4**, **5**) or heterozygous (GGGGC)_{4/5} (**4/5**) standards were separated by PAGE. Upper (~200 bp) and lower (~170 bp) bands of the heterozygous (GGGGC)_{4/5} repeat PCR product were excised from the gel, and the DNA was reamplified (**Up** and **Lo**, respectively). PCR product of homozygous (GGGGC)₄ and ₅ repeats, in equal amounts, were combined, melted, and cooled (**4+5**). Lane 1 is a no template control (**0**). **D**) All possible homozygous or heterozygous combinations of the polymorphism from constructs containing **2**, **3**, **4**, **5**, and **6** numbers of repeats, as indicated, were amplified and separated by PAGE. The 100-bp ladder is indicated (**L**).

Figure 4.

AHR promoter polymorphisms in tumor and non-tumor tissue. DNA from cancer patients (numbers **1 – 16**) using either non-tumor (**N**) or tumor (**T**) tissue sources was amplified and separated on polyacrylamide gels for heteroduplex analysis. Primary cancers were **A) 1**, gastric; **2 – 8**, lung; **9**, tongue; **10**, peritoneal; and **B) 11 – 16**, colon. Electrophoretic standards for (GGGGC)₄ or 5 (**4, 5**), (GGGGC)_{2/4} (**2/4**), and (GGGGC)_{4/5} (**4/5**) are boxed. The no-template controls are labeled (**0**), and the 100-bp ladder is indicated (**L**).

Figure 5. *AHR* promoter polymorphisms in tumor tissue. DNA from cancer patients (numbers 1 – 29) using tumor tissue sources was amplified and separated on polyacrylamide gels for heteroduplex analysis. Primary cancers were A) 1, esophagus; 2 – 3, sarcoma; 4, ovary; 5, lung; 6, bladder; 7, gastric; 8 – 9, endometrium; 10 – 14, breast; 15 – 21, colon; and B) 22 – 29, breast. Electrophoretic standards for (GGGGC)₄ or 5 (4, 5), (GGGGC)_{2/4} (2/4), and (GGGGC)_{4/5} (4/5) are boxed. The no-template controls are labeled (0), and the 100-bp ladder is indicated (L).

Figure 6.

Allele frequencies of the repeat polymorphism in cancer patients and NYS neonates. Represented are the proportion of alleles containing (GGGGC)₂, (GGGGC)₄, (GGGGC)₅, and (GGGGC)₆, as indicated, and (GGGGC)₂GGGGAGGGGC (indicated as **C/A SNP**) in NYS neonates (black bars), lung cancer patients (white bars), colon cancer patients (medium gray bars), and breast cancer patients (hatched bars). Broken brackets represent *P*-values for comparison across all sample groups. Solid brackets represent *P*-values for pair-wise sample group comparisons with *P*<0.05. For each sequence repeat polymorphism, a test *P*-value of approximately 0.007 corresponds to a true type-1 error rate of 0.05 (i.e., *P*=0.05 divided by 7 tests per locus).

Summary of Individuals with repeat polymorphism containing an SNP

Subject	Diagnostic Group	Genotype of Repeat	SNP	Gender	Ethnicity	Diagnosis
1	Neonate	(GGGGC) ₄	C/A *	Male	African-American	NA
2	Neonate	(GGGGC) ₄	C/A *	Male	Hispanic	NA
3	Neonate	(GGGGC) ₄	C/A *	Male	Hispanic	NA
4	Neonate	(GGGGC) ₄	C/A *	Female	Hispanic	NA
5	Neonate	(GGGGC) _{4/6}	G/T **	Female	Caucasian	NA
6	Lung Cancer	(GGGGC) ₄	C → A ***	Female	African-American	History of lung cancer; Moderately differentiated Adenocarcinoma consistent with new primary tumor; emphysema, no invasion to pleura
7	Colon Cancer	(GGGGC) ₄	C/A *	Male	African-American	Metastatic colon cancer to liver; adenocarcinoma

C→A: GGGCGGGCGGGG(A/A)GGGGC

Table 2

Distribution of race/ethnicity and gender among diagnostic groups.

Subjects	Caucasian	Hispanic	American African	Asian	Unknown	Ethnicity	Male	Female	Unknown Gender	Total
Neonates	54	32	32	32	0		77	73	0	150
Lung Cancer	60	3	23	1	7		53	41	0	94
Colon Cancer	36	0	3	1	4		24	19	1	44
Breast Cancer	15	0	10	1	13		0	36	3	39

Table 3

Distribution of diagnoses among disease groups.

Cancer	Number of Individuals, Diagnosis
Lung	50, adenocarcinoma; 25, squamous cell carcinoma; 2, carcinoid; 4, bronchoalveolar; 2 small cell carcinoma; 5, non-small cell carcinoma; 3, large cell carcinoma; 1, mesothelioma; 2, mixed bronchoalveolar/adenocarcinoma
Colon	42, adenocarcinoma; 2, unspecified
Breast	19, ductal carcinoma; 4, lobular carcinoma; 2, both ductal and lobular features; 14 unspecified. Of these, 11 were specified as adenocarcinoma

Table 4

Primate DNA analysis of the (GGGGC)_n repeat. Deviations from the repeat are shown in bold.

Species	Common Name	Number of Individuals Sequenced	Sequences of Region Homologous to human (GGGGC) _n repeat	Distance from TSS
<i>Macaca nemestrina</i>	Pig-tailed macaque	8	(GGGGC) ₂	not determined
<i>Macaca fascicularis</i>	Cynomolgus macaque	8	(GGGGC) ₂	not determined
<i>Macaca mulatta</i>	Rhesus macaque	8	(GGGGC) ₂	-11,306*
<i>Gorilla gorilla</i>	Western gorilla	6	(GGGGC) ₄ (GGGGC) ₆ (GGGGC) ₂ (GGGGC) ₅ (GGGGC) ₇ (GGGGC) ₈	-295**
<i>Pan troglodytes</i>	Chimpanzee	8	(GGGGA)(GGGGC) ₂	-89**
<i>Homo sapiens</i>	Human	327	(GGGGC) ₄ (GGGGC) ₅ (GGGGC) ₂ (GGGGC) ₆ (GGGGC) ₂ (GGGGA)(GGGGC)	-28
<i>Pan paniscus</i>	bonobo	****	(GGGGA)(GGGGC) ₂	-58**
<i>Nomascus leucogenys</i>	gibbon	***	(GGGGC) ₂	-11,394*
<i>Pongo abelii</i>	orangutan	***	(GGGGC) ₂	+152
<i>Callithrix jacchus</i>	marmoset	***	(GGGGC) ₂	+63

* TSS is present at a region homologous to the beginning of exon 2 of the human *AHR*.

** The TSS in the bonobo, chimpanzee, and gorilla results in a loss of about 5%, 10% and 40% of the 5'-untranslated region of the mRNA, respectively, in comparison with that of the human mRNA.

*** Sequence from Ensembl database.

**** Sequence of bonobo: <http://www.ncbi.nlm.nih.gov/gene/?term=AHR%2C+bonobo> (Maglott et al., 2005)