

***AHR* gene variants in patients with cancer and their presence in the Latin American population.**

NA. Trujillo-Pelayo¹, S. Guauque-Olarte², Cl. Vargas Castellanos¹, L. Cifuentes-C³

¹Human Genetics Research Group- GENEHUIS. Universidad Industrial de Santander, Bucaramanga-Colombia

²GIOM Research Group, Universidad Cooperativa de Colombia, Envigado-Colombia

³GIOD Research Group, Universidad Cooperativa de Colombia, Pasto-Colombia

Abstract

The aryl hydrocarbon receptor (AhR), encoded by the *AHR* gene, is a transcription factor belonging to the basic helix-loop-helix (bHLH) Per-Arnt-Sim (PAS) family. The AhR plays a crucial role in the regulation of xenobiotic metabolism, cellular differentiation, stem cell maintenance, immunity and cancer. The activation of the canonical AhR signaling pathway inducing downstream genes, such as cytochrome P450 enzymes, involved in the detoxification metabolism in the human body, plus the interaction of the Ah Receptor with several non canonical signaling pathways, like the epidermal growth factor receptor (EGFR), signal transducer and activator of transcription 3 (STAT3), hypoxia-inducible factor-1 α (HIF-1 α), nuclear factor κ B (NF- κ B), estrogen receptor (ER), and androgen receptor (AR) has recently gained

significant attention as potential targets for the development of novel cancer therapies. AHR genetic variants have been reported in a variety of cancers including, glioblastoma, ovarian, colorectal, gastric, lung, skin, and head and neck cancer in studies performed mainly in White populations. Here, we present a compilation of genetic variants in the *AHR* gene identified in cancer genomic studies in different populations, with specific focus on the Latin American population.

We performed a search for variants in the *AHR* gene in three cancer data repositories (COSMIC, ICGC and TCGA), and two genomic data repositories (GnomAD v2.1 and The GWAS Catalog). Additionally, *in silico* analysis of variant pathogenicity was performed using bioinformatics tools.

This study found a high number of variants in *AHR* in different cancer types, with a total of 64 potentially impactful variants to the protein function. These genomic changes were reported in studies from United States, United Kingdom; China, among other countries. The cancer types with most high impact variants in *AHR* were Colorectal, Ovary, Stomach, and brain cancer. The identification of *AHR* variants in Latin American populations was established mostly through the ICGC and TCGA repositories and 7 genomic alterations were detected as high impact consequence assuming to have a disruptive impact in the protein. Latin American studies have been restricted mostly to Skin, Uterus and Cervix cancer. Furthermore, the observed distribution of variants including, insertion-deletions, frameshift and stop gained, in the protein domains: bHLH, PAS A and PAS B may suggests a weakening of the formation of a stable AhR-ARNT heterodimer necessary for the AhR signaling pathway. Also, 41 high impact alterations were found covering the genomic region around exons 7–9 in AhR (Chr7:17372389–17378532), which may

lead to a constitutively activated AhR signaling inducing the promotion of cancer. This result may correlate with the variety of cancer types where *AHR* gene was expressed.

This study presents a compilation of genetic variants in *AHR* with high impact in several cancer sites reported in White, Europe, Hispanic and other populations. The underrepresentation of Latin Americans found in these genomic repositories indicates the need to study the *AHR* variation in this population allowing to identify genetic risk variants in the gene that leads to a better understanding of the receptor's role in the development and progression of cancer.

Keywords: *AHR* gene, cancer, Latin America, genomic databases, variants.

Introduction

Cancer is the second leading cause of death worldwide behind cardiovascular diseases and its incidence and mortality are rapidly growing worldwide (Bray et al. 2018, Sung et al. 2021). According to the International Agency for Research on Cancer, a global estimate of 19.3 million new cancer cases and almost 10.0 million cancer deaths occurred in 2020 (Sung et al. 2021); this reflects changes in the prevalence and distribution of major risk factors for cancer like tobacco and alcohol consumption, poor diets, infectious agents, environmental pollutants, and radiation exposure (Yang et al. 2019). Although environmental exposures to carcinogens contribute to cancer risk, variation in the incidence and progression of cancers among individuals can be attributed to interindividual variation (Wilson et al. 2002; Yang et al. 2019).

The Aryl hydrocarbon receptor (AhR) a protein encoded by the *AHR* gene, is a member of the basic-helix-loop-helix (bHLH)-Per-ARNT-Sim (PAS) family of transcription factors (Powell et al. 2013; Hanieh et al. 2016; Baker et al. 2019). As a member of the bHLH superfamily, AhR contains a bHLH, a Per-ARNT-Sim (PAS), and a glutamate-rich domain (Hu et al. 2023). The bHLH domain is situated at the N-terminus of AhR, and it connects the receptor to the promoter region of its target genes (Hu et al. 2023). The PAS domain, which can be divided into PAS-A and PAS-B domains, is located between the bHLH and the glutamate-rich domains (Hu et al. 2023). Both the bHLH domain and the PAS-A domain of AhR are involved in the heterodimerization process with the aryl hydrocarbon receptor nuclear translocator

(ARNT), while the PAS-B domain mainly connects AhR to its ligand and helps in AhR nuclear translocation. The PAS-B domain also serves as the ligand binding domain (LBD) (Hu et al., 2023). The glutamate-rich domain, also known as the transcriptional activation domain (TAD), is located at the C-terminal of the AhR protein. It plays an important role in recruiting transcriptional activators once the AhR-ARNT heterodimer has bound to the xenobiotic response element (XRE) (Hu et al., 2023).

The activation of the canonical AhR signaling pathway is widely recognized for its role in driving the expression of specific targets that are involved in the regulation of drug metabolism and detoxification. This pathway is responsible for the induction of cytochrome P450 enzymes, which play a crucial role in the metabolism of toxins within the human body (Sato et al., 2019). Chemicals present in tobacco have been demonstrated to enhance AhR expression and activities, leading to the biosynthesis of CYP1A1 and CYP2E1. These phase I detoxification enzymes are crucial in eliminating harmful tobacco chemicals from the human body. Besides, cytochrome P450 enzymes like CYP1A1 and CYP1B1 activate pre-carcinogens by the formation of DNA adducts in human cells (Alsubait et al., 2020). Apart from the canonical signaling pathway, AhR could also interact with various proteins from different signaling pathways (Mulero and Fernandez 2016; Roman et al. 2018; Lafleur et al. 2023) such as the expression of epidermal growth factor receptor (EGFR) to activate the phosphoinositide 3-kinase/protein kinase B (PI3K/AKT) signaling pathway (Popolo et al. 2017; Bai et al., 2023), the FAK/Src (Tomkiewicz et al. 2013; Wei et al. 2018), TGF- β (Gagliani et al. 2015), and NF- κ B signaling pathways (Galle-Treger

et al. 2016) critical to cell normal homeostasis (Feng, Cao y Wang, 2013). Also, AhR could form a heterodimer with the signal transducer and activator of transcription 3 (STAT3) to inhibit STAT3-mediated signaling activities (Nukaya et al., 2004; Sondermann et al., 2023). Additional functions have been associated with the Ah receptor relating to apoptosis, gene regulation, cell proliferation and migration (Feng, Cao y Wang, 2013; Yang et al. 2019), normal physiology and embryonic development (Bradshaw and Bell 2009; Stejskalova y Pavek, 2011; Murray, Patterson y Perdew, 2014; Yang et al. 2019) and the correct functioning of the female reproductive system at all stages from fetal development through adulthood (Baker et al. 2019; Karman et al. 2009).

As AHR is known to be activated by various endogenous and exogenous ligands and to form various interactions with other signaling proteins, dysregulation in AhR expression and activity is likely to promote cancer development (Feng, Cao y Wang, 2013; Gearhart-Serna et al. 2020; Benoit et al. 2023), thus the AhR has emerged as an important novel drug target for cancers by enhancing or inhibiting carcinogenesis in a wide range of tissues (Safe, Cheng y Jin, 2017). Depending on the malignancy type, glioblastoma, ovarian, colorectal, gastric, lung, skin, head and neck cancer and the endogenous role of the receptor, clinical treatment with AhR agonists or antagonists alone or in combination with other drugs constitutes an approach for cancer (Safe, Cheng y Jin, 2017). The difference in transcriptional activity of the Ah receptor linked to polymorphic variation in the *AHR* gene may contribute to individual genetic cancer susceptibility (Miller et al, 2002; Norppa, 2003; Chen et al. 2009; Brunotto et al. 2014; Fernández et al. 2019; Re et al. 2020; Cheng et al. 2022).

However, most genomic studies in cancer include individuals from The United States and Europe (Carrio et al. 2020; Chong et al. 2023, p. 8-10), while Latin American population continue to be underrepresented, which limits the findings from genomic research to clinical care in diverse populations (Popejoy & Fullerton, 2016; Landy et al. 2018). Reference genomics databases are a powerful tool for understanding the biological function of genetic variation. Their use guides the identification of pathogenic variants and supports the analysis of disease-causal/susceptible genes relationships (Gudmundsson et al. 2022). Accordingly, in the present study we gathered data from three cancer and two genomic repositories: The Catalogue of Somatic Variants in Cancer-COSMIC (<https://cancer.sanger.ac.uk/cosmic>), which is the largest database of somatic mutations and their effects on human cancer (Forbes et al. 2015); The International Cancer Genome Consortium-ICGC (<http://icgc.org/>), that is a global data portal containing a large catalog of mutational abnormalities in the major tumor types (Zhang et al. 2019); The Cancer Genome Atlas-TCGA (<https://portal.gdc.cancer.gov/>), a project covering a large cancer-causing genomic alterations and clinical data across different tumor types (Tomczak, Czerwińska y Wiznerowicz, 2015); The Genome Aggregation Database-GnomAD v2.1 (<https://gnomad.broadinstitute.org/>), which is currently the largest collection of population variation from harmonized exome and genome sequencing data from around the world (Gudmundsson et al. 2021) and The GWAS catalog (<https://www.ebi.ac.uk/gwas/>). All variants in COSMIC, ICGC and TCGA contained confirmed somatic alterations across tumor types. The GnomAD “cancer” dataset and some variants reported in the GWAS Catalog were identified from blood

(germline) cancer samples. The objective of the present study was to compile genetic variants in the *AHR* gene identified in cancer genomic studies in different populations, with specific focus on the Latin American population, and to determine the classification of the severity of the variant's consequence on the receptor based on agreement with *in silico* analysis tools. The Identification of *AHR* cancer risk variants may increase our comprehension of the role of the receptor in development and cancer progression, and provide clues for early tumor detection and the development of targeted cancer drugs.

Methods

Data consolidation from cancer genomics data repositories and the GWAS catalog

Genetic variants in *AHR* were downloaded from three cancer data repositories: COSMIC, ICGC, and TCGA; and two genomic data portals: GnomAD v2.1 and the GWAS catalog, last accessed September 30, 2022. For the GnomAD repository, the search was broken down into three steps (1) extraction of variants from the whole GnomAD v.2.1.1 dataset; (2) extraction of variants from the GnomAD v.2.1.1 (non-Cancer) dataset; (3) a subtraction of the variants in the GnomAD (non-Cancer) dataset from the whole Gnomad dataset to obtained a list of variants from individuals with cancer; this new dataset was called GnomAD "Cancer". For the GWAS Catalog, only variants in the *AHR* gene from cancer studies were extracted. Then, the list of variants from all repositories was consolidated in a single data file. Duplicate variants

were removed filtering by “*Genomic DNA Change*” keeping the information about which databases contain the variant (Figure 1). Data obtained from cell line studies were not included.

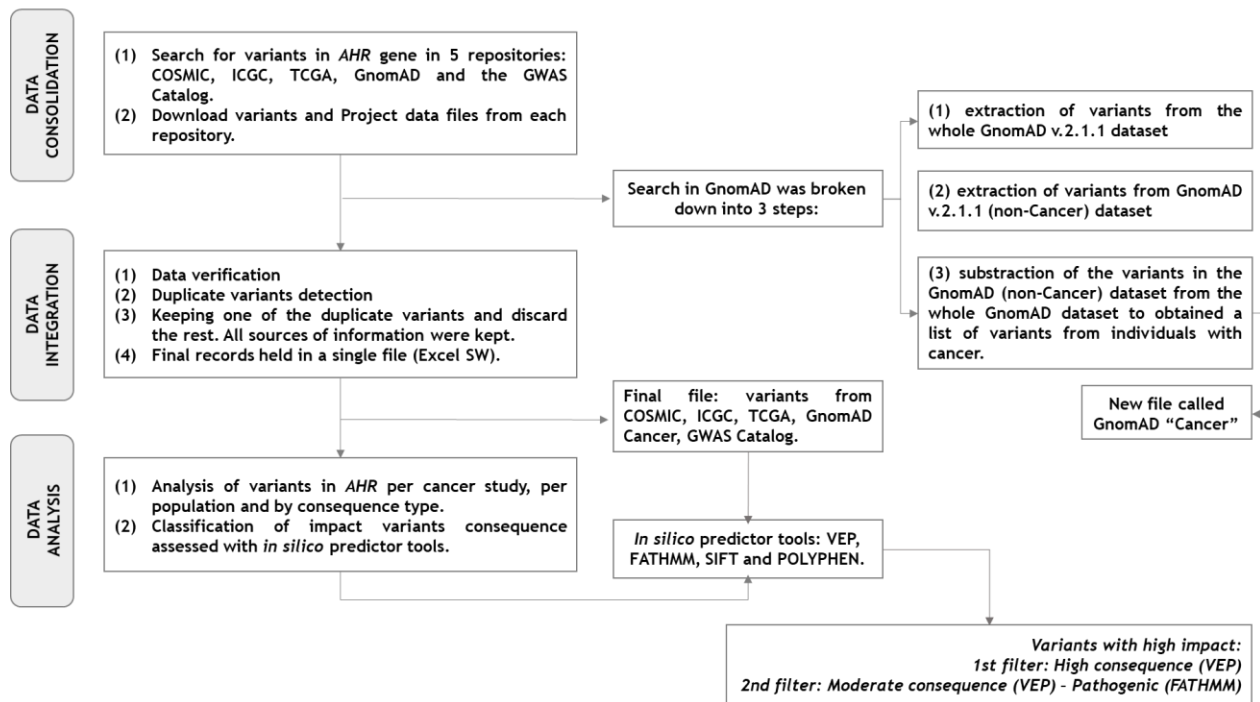


FIGURE 1. Flow chart showing steps for data consolidation from cancer genomics repositories. COSMIC, ICGC, TCGA, GnomAD and the GWAS Catalog.

Functional impact of the variants

Each genomic database uses different tools to predict the consequences of the variants on the protein sequence assigning functional impact scores. COSMIC employs FATHMM (Shihab et al. 2013) and the data provided for each variant is a score and a qualitative prediction (either “pathogenic”, “neutral” or “benign”). ICGC consolidate predictions from Mutation Assessor (Reva, Antipin, & Sander, 2007; 2011), SIFT (Kumar, Henikoff & Ng, 2009; Vaser et al. 2016) and FATHMM, and

assign the labels “unknown”, “low”, “medium” or “high”. TCGA uses the Ensembl Variant Effect Predictor (VEP) (McLaren et al. 2016; Hunt et al. 2021) which displays a classification of the severity of the variant consequence corresponding to “modifier”, “low”, “moderate” or “high” impact. GnomAD uses two in silico tools: SIFT and Polyphen (Adzhubei et al. 2010; Adzhubei, Jordan, & Sunyaev, 2013) and for each amino acid substitution an impact score with a qualitative prediction (“probably damaging”, “possibly damaging”, “benign” or “unknown”) is provided. Since SIFT and PolyPhen only predicts whether an amino acid substitution affects protein function (coding non-synonymous variants), and considering that impact score for each variant was not available in all repositories, we run the predictors FATHMM and VEP manually to predict and analyze the functional impact of all variants contained in the final dataset.

To detect variants in *AHR* with pathogenic potential, the dataset was filtered to extract: (1) Variants with VEP predictions of high and (2) Variants with VEP predictions of moderate and pathogenic label by FATHMM.

Data extraction covered from databases included: source, genome position, rs number or existing variation, primary site, reference and alternate allele, protein change, consequence, impact scores, germline or somatic mutation status and population. The frequencies of variants in *AHR* per cancer study (primary site reported by project), by consequence type (according to functional role) and per populations were calculated. *AHR* variants in the Latin American population were examined deeply. Data visualization was performed using InteractiVenn to create

Venn diagrams, (Heberle et al, 2015) and R v.4.0.0 (R Core Team, 2019) over R Studio v.1.3. 1073 (RStudio Team, 2019).

Results

Results per repository

Cancer Data Repositories

The Catalogue of Somatic Mutations in Cancer-COSMIC

The Catalogue of Somatic Variants in Cancer (COSMIC) is the largest resource for exploring the impact of somatic mutations in human cancer. It has integrated 736 key cancer genes including 49 tissue type from 1,515,965 samples (latest Census: COSMIC v97, Nov 2022). For the purpose, COSMIC database reported 1,301 genetic variants in the *AHR* gene in human cancer tissues. Primary Liver malignancy was the most common tumor site including 15% of the total number of variants followed by Breast (14%) and Pancreas (9%) cancers. This dataset contained mostly intronic (73%) and protein coding (20%) variants. Among coding mutations, genetic consequences included missense (76%) and synonymous (24%) variants.

The 76% of the variants were classified with modifier impact, 5% with low impact, 16% as moderate and 3% as high. According to VEP, from the 37 variants of high impact, 18% corresponded to Large Intestine cohort studies, followed by 11% variants present in Stomach and 10% in Lung cohort analysis. Primary tumors of the Endometrium and Kidney reported the lowest number of high impact variants

equivalent to 3% each. In regard to consequence variant type, annotations were observed as follows: Stop-gained (59%), Frameshift (24%), Acceptor splice site (8%), Missense (3%), Synonymous (3%) and Donor splice site (3%).

A total of 133 additional variants had VEP moderate impact and FATHMM pathogenic label and 97% of them corresponded to missense variants. Most mutations were observed in Skin (16%), Large Intestine (13%) and Lung (13%) cohort studies.

International Cancer Genome Consortium-ICGC

The International Cancer Genome Consortium (ICGC) data portal contains information about 57,905 mutated genes and 21 tumor sites from 24,289 sample donors (latest data release 28, Nov 2019). From our search in ICGC a total of 1,124 genetic variants in *AHR* gene were identified amongst 65 worldwide cancer projects. Most variants were found in Skin (16%), Liver (15%) and Esophagus (9%) cancers, which corresponded to the most common malignancies contributing to 13%, 15% and 10% of the total number of cases informed. Variants contained in the database mostly included intronic genetic changes (51%), Missense (15%) and Upstream (9%) variants, whereas Inframe deletion, Acceptor splice site and Start-lost variants were the less (0.27%) represented genetic consequences of the dataset.

With regard to potential effects of genetic variants we assessed *in silico* their functional impact with VEP. Most variants had modifier impact (77%), follow by 5% with low (5%) impact, moderate with 15% and high with 3%. Overall, we observed

27 high-impact variants in *AHR* gene in patients with Bladder (19%), Esophagus (11%) and Ovary (11%) cohort studies and at least one pathogenic variant was identified in Breast, Colorectal, Kidney, Skin and Uterus projects. About 81% of the variants found coincide to Stop-gained genetic consequences and most of them have been reported in populations from The United States (63%), China (15%) and United Kingdom (11%).

Among the variants with VEP moderate and FATHMM pathogenic impact, 101 missense variants (9%) were observed in the dataset. The primary cancer sites with a greater number of variants were as follows: Liver (16%), Skin (14%), Lung (12%) and Bladder (11%), mostly reported in American (The United States, 69%) and Chinese (China, 17%) populations.

We further explored the entire dataset obtained from ICGC portal and we found 21% *AHR* variants reported in American population, but we only detected 2% genetic changes in the Latin American population represented by the Skin cohort study in Brazil with 19 cases and 27 variants reported in the project. None of these variants in Latin Americans corresponded to a high functional impact in the protein product and the most frequent variants observed in this population were chr7:g.17389450->T and chr7:g.17375399G>C.

The Cancer Genome Atlas-TCGA

The Cancer Genome Atlas (TCGA) is a database covering 22,369 genes and 67 primary sites from 85,513 cases. For this study, TCGA provided a dataset containing

229 variants in *AHR* from 28 projects. All studies only included cancer samples. This allowed us to identify primary malignancies mainly in Ovary (18%), Bronchus and lung (14%), Breast (10%) and Corpus Uterus (10%); Also, Corpus uterus (28%), Bronchus and lung (11%), and Bladder (10%) displayed the higher number of variants in *AHR* gene. Analysis of genetic consequences determined 52% missense, 17% synonymous and 16% 3' prime UTR variants, whereas the less represented genetic change was splice region with only one variant reported.

Further analysis using VEP, allowed us to detect 26 variants (11%) with a high functional impact from the entire dataset. From these, 54% and 46% corresponded to Frameshift and Stop-gained variants. Most variants were found in Stomach and Bladder cohort studies with 5 and 4 variants. We noticed that only 4 high variants (15%) have been reported in White-Hispanic or Latino population, and the most frequent variant found was chr7:g.17335775G>C.

Among the variants with VEP MODERATE and FATHMM "PATHOGENIC" impact, a total of 71 missense variants were observed in the data. In accordance with primary malignancy sites, most variants were reported in Uterus (13 variants – 18.3%), Bladder (12 variants – 17%), Bronchus and lung (12 variants – 17%), and Skin (11 variants – 15.5%). From the entire TCGA dataset we recognized 11 variants in Latin American population, where two cases from the Corpus uteri cohort presented the higher number of variants in the *AHR* gene (5 variants – 45.4%).

Genomic Data Repositories

GnomAD non-cancer

The Genome Aggregation Database (GnomAD) v2.1 (<https://gnomad.broadinstitute.org/>) has aggregated 3,060 genes, 15,708 whole genomes and 125,748 exomes from 195,000 individuals (Gudmundsson et al. 2021). According to our search, GnomAD (v.2.1.1) database displayed a total of 4,359 *AHR* variants. To identify only variants present in cancer patients, we performed a second search using the dataset named GnomAD v.2.1.1 Non Cancer, which allowed us to detect 4,332 *AHR* variants. After finding and removing duplicates from the two data sets, GnomAD and GnomAD Non Cancer, 29 unique cancer risk variants were identified in the data file and all of them corresponded to germline variants.

The dataset containing 29 Cancer variants, mostly included missense (55%) and intronic (21%) variants, whereas the genetic consequences less represented concerned frameshift (1) and 5' UTR (1) variants, which represented about 7% of the total variants identified.

Annotation and filtering of genomic variants using VEP permitted us to detect variants with high and moderate impact, which represented about 59% of all variants observed in the dataset. Categorically, only 1 high impact variant was identified and corresponded to a Frameshift alteration. Variants in the GnomAD Cancer dataset have been reported in European Non Finnish (23 variants – 79%), African (3 variants – 10%), South Asian (2 variants – 7%), Latino (1 variant – 3%), Ashkenazi Jewish (1 variant – 3%) and East Asian (1 variant – 3%) populations. From these, the

rs980233827 and rs771303532 were observed in both African and European (Non-Finnish) populations.

Among the variants with VEP moderate and FATHMM pathogenic impact, 4 genomic changes were identified corresponding to missense variants, and have been informed in Europeans Non Finnish (50%), Hispanics (25%) and South Asian (25%) populations. In further analysis considering only Latin American population, we detected 1 *AHR* variant in the entire GnomAD Cancer dataset and corresponded to a possibly pathogenic alteration with a moderate and pathogenic functional impact according to VEP and FATHMM (rs779613705, missense variant).

Genome-Wide Association Studies (GWAS) Catalog

The GWAS catalog (<https://www.ebi.ac.uk/gwas/>) is a collection of all human genome-wide association studies containing about 45,000 published GWAS across more than 5000 human traits, including cancer (Sollis et al. 2023). Search for variants using GWAS catalog allowed us to identify 6 *AHR* genomic variant studies reporting a total of 4 unique variants as follows: chr7:g.16944656C>T, chr7:g.17014722T>C, chr7:g.17095084G>A and chr7:g.17203933T>C. All variants were intronic with non functional impact of amino-acid substitutions in the protein. These variants have been reported in several primary cancer sites (Skin, Lung and Blood) in different ancestry populations, European, African American and Hispanics (Dataset 1-Supplementary material 1).

Data integration from repositories

Variants in AHR and their frequency

After consolidating the variants identified in the cancer and genomic repositories, overall, 7,019 variants in the *AHR* gene were identified: 1,301 were reported in COSMIC, mostly somatic alterations and just a few variants informed as “unknown origin”, 1,124 somatic variants in ICGC, 229 somatic variants in TCGA, 4,359 germline variants in GnomAD, and 6 variants in the GWAS catalog, some reported as germline variation and some other as alterations of unknown origin.

From this number 1,628 variants were reported by a sole repository (879 in COSMIC, 676 in ICGC, 41 in TCGA, 28 in GnomAD “Cancer” and 4 in the GWAS Catalog), and 451 were found in at least two databases, for a total of 2,079 variants identified in the *AHR* gene. Among the three cancer repositories COSMIC, ICGC and TCGA, without duplicates, 1,596 possibly pathogenic variants in *AHR* were observed (Figure 2). A complete list of variants identified in *AHR* gene is available in the Dataset 1-Supplementary material 1.

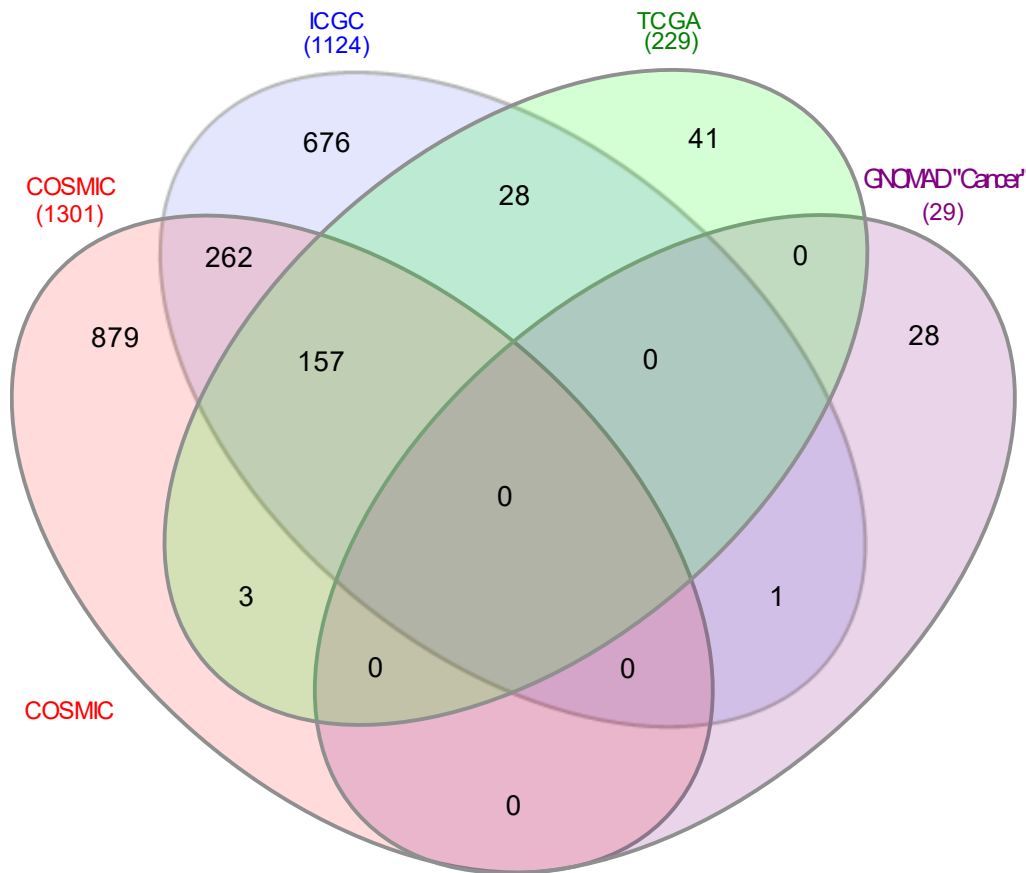


FIGURE 2. Distribution of variants in *AHR* gene. Venn diagram representing the total number of variants reported in each database (COSMIC, ICGC, TCGA, GNOMAD "Cancer") after removing duplicates. Additionally, 4 variants were reported in the GWAS catalog (not shown). DOI: 10.1186/s12859-015-0611-3

The primary sites with the highest number of variants in *AHR* were Liver (314 variants – 15%) and Breast (229 variants – 11%) and the less represented were Lymph nodes, Thymus, and Testis (Figure 3).

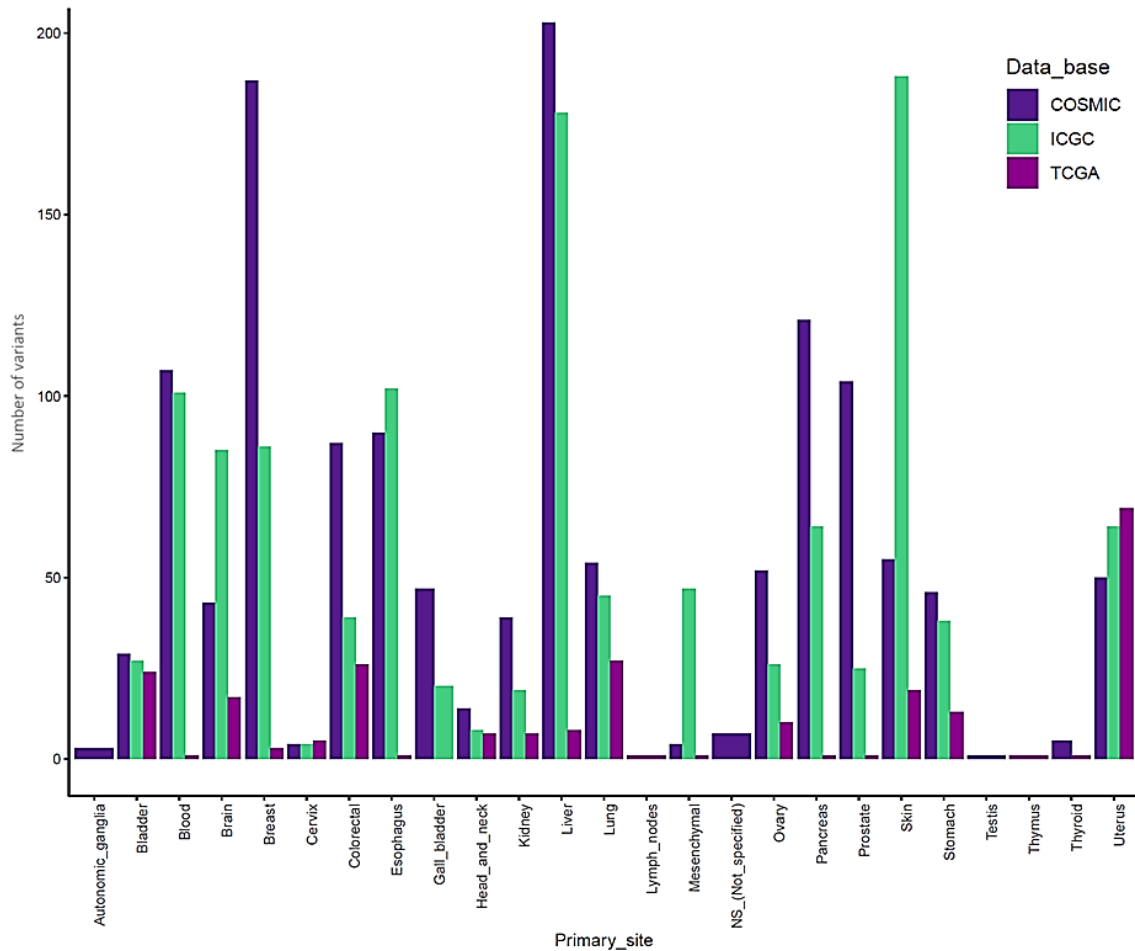


Figure 3. Primary sites reported in *AHR* gene per cancer study. The X axis shows the type of cancer in which at least one mutation in *AHR* was identified, the Y axis indicates the number of variants reported for each cancer type. COSMIC shown in indigo reported a total of 1,301 *AHR* variants, ICGC shown in green reported a total of 1,124 alterations and TCGA shown in purple reported a total of 230 genomic changes in the gene.

Classification of variants in *AHR* and their impact

Variants were grouped according to consequence, intronic variants were the most frequent (64.6%) followed by missense (13.3%), 3' prime UTR (5.1%) and upstream

variants (4.9%). Splice region, inframe deletion or insertion and stop-gained had a frequency lower than 2.0% (Figure 4).

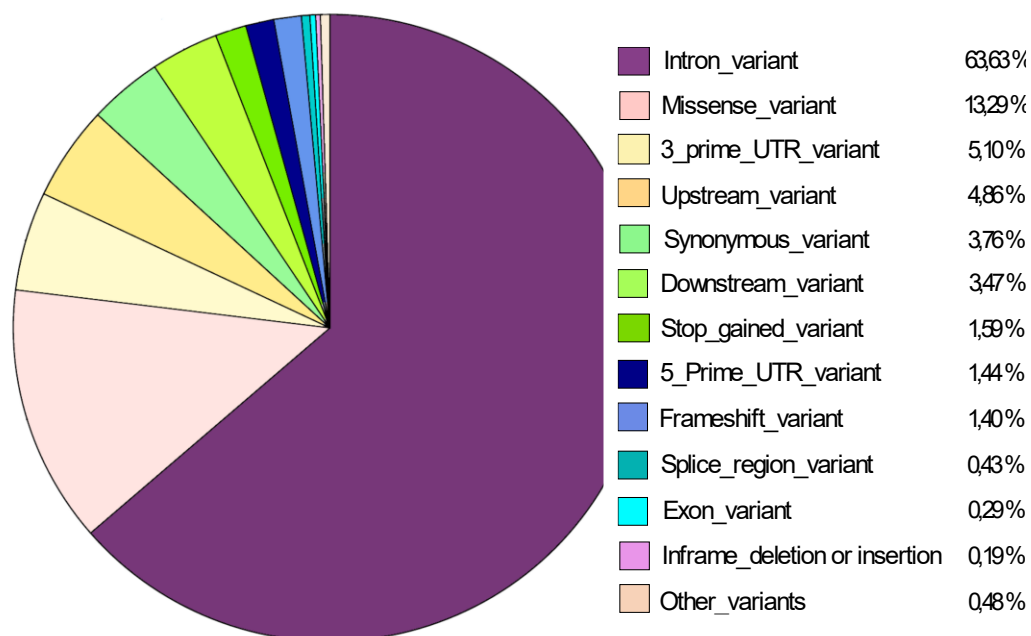


FIGURE 4. Distribution of variant consequence. Pie chart showing the percentage distribution of each variant consequence reported in the *AHR* gene from COSMIC, ICGC, TCGA, GnomAD “Cancer” and GWAS Catalog repositories.

Among the 2,079 variants identified, 64 (3.1%) were predicted to have high impact according to VEP, including truncating, insertion and deletion (INDEL) and start lost variants (Figure 5). A total of 11 truncating variants, out of the 32 stop-gained variants found, occurred in the PAS-A and PAS-B domain and were more frequently observed in Brain cancer. In contrast, 13 insertion-deletion variants, out of 27 frameshift alterations, occurred in the transactivation domain (TAD) and were more

frequently observed in Ovary and Colorectal cancer (Figure 5; Table 1). The most frequent high-impact variant identified in the dataset was chr7:g.17339390delC (Table 1).

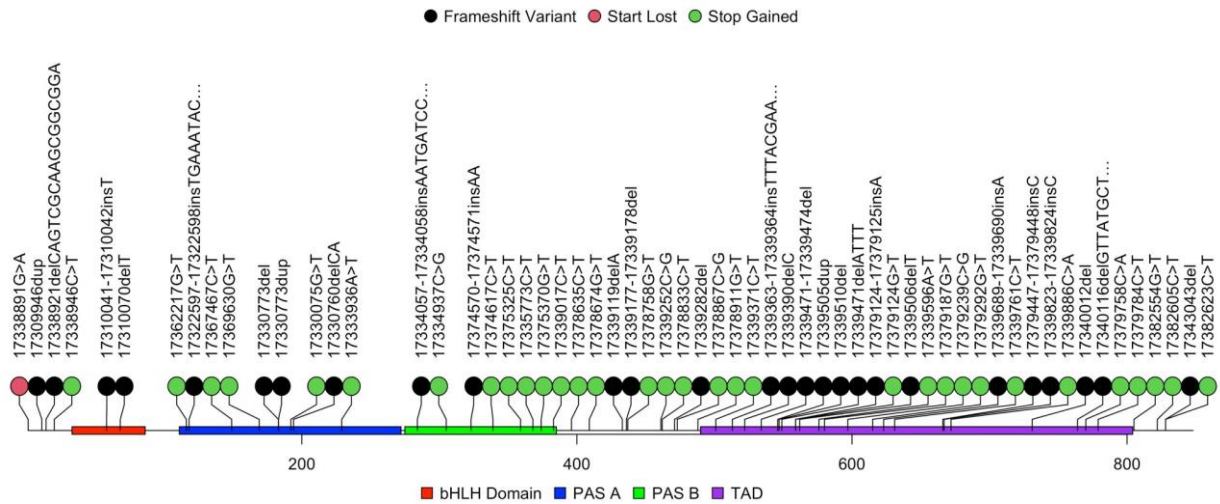


FIGURE 5. Schematic representation of *AHR* variant distribution along the gene. The 64 *AHR* variants from patients reported from cancer genomic studies in the COSMIC database, ICGC, TCGA, and GnomAD Cancer predicted to have high impact according to VEP. bHLH domain, basic helix loop helix domain; PAS-A and PAS-B domain; TAD, Transcription activation domain.

Additionally, 155 variants (7.4%) were predicted to have both moderate and pathogenic impact by VEP and FATHMM, respectively. The most frequent variants found were chr7:g.17378642G>A and chr7:g.17375399G>C (Dataset 2-Supplementary material 2). The identification of all *AHR* variants among reported genomic repositories is available in Dataset 3-Supplementary material 3.

Table 1. Variants in *AHR* with high impact. A total of 64 variants were predicted by VEP to have high impact. The variants identified in the GWAS catalog were intronic and do not have high impact on the protein.

| SOURCE | PRIMARY SITE | GENOMIC DNA CHANGE | PROTEIN CHANGE | DONORS AFFECTED | SOMATIC OR GERMLINE STATUS |
|------------------|-----------------------------|-------------------------------|----------------|------------------|----------------------------|
| ICGC | Esophagus | chr7:g.17374570_17374571insAA | Q323Q? | ESAD-UK: 1/409 | Somatic variant |
| ICGC | Ovary | chr7:g.17379124_17379125insA | E559R? | OV-US: 1/426 | Somatic variant |
| ICGC | Kidney | chr7:g.17379447_17379448insC | Q666H? | KIRC-US: 1/361 | Somatic variant |
| ICGC COSMIC | Skin | chr7:g.17382554G>T | G805* | SKCM-US: 1/466 | Somatic variant |
| ICGC TCGA COSMIC | Stomach | chr7:g.17379187G>T | E580* | STAD-US: 1/439 | Somatic variant |
| ICGC TCGA COSMIC | Blood | chr7:g.17338946C>T | Q20* | DLBC-US: 1/38 | Somatic variant |
| ICGC | Blood | chr7:g.17338891G>A | M1I | MALY-DE: 1/241 | Somatic variant |
| ICGC COSMIC | Bladder | chr7:g.17379239C>G | S597* | BLCA-CN: 1/103 | Somatic variant |
| ICGC | Brain | chr7:g.17375325C>T | R359* | PEME-CA: 1/112 | Somatic variant |
| ICGC TCGA COSMIC | Bladder | chr7:g.17382623C>T | Q828* | BLCA-US: 1/411 | Somatic variant |
| ICGC | Liver | chr7:g.17378674G>T | G409* | LICA-CN: 1/402 | Somatic variant |
| ICGC COSMIC | Esophagus | chr7:g.17379292G>T | E615* | ESAD-UK: 1/409 | Somatic variant |
| ICGC TCGA | Brain | chr7:g.17379758C>A | S770* | GBM-US: 1/388 | Somatic variant |
| ICGC COSMIC | Colorectal | chr7:g.17375267A>T | p.? | COAD-US: 1/402 | Somatic variant |
| ICGC TCGA COSMIC | Stomach | chr7:g.17378635C>T | R396* | STAD-US: 1/439 | Somatic variant |
| ICGC | Liver | chr7:g.17369630G>T | E169* | LICA-CN: 1/402 | Somatic variant |
| ICGC COSMIC | Breast | chr7:g.17378867C>G | S473* | BRCA-UK: 1/141 | Somatic variant |
| ICGC TCGA COSMIC | Bladder | chr7:g.17367467C>T | Q149* | BLCA-US: 1/411 | Somatic variant |
| ICGC COSMIC | Ovary | chr7:g.17378758G>T | G437* | OV-US: 1/426 | Somatic variant |
| ICGC TCGA COSMIC | Uterus | chr7:g.17362217G>T | E116* | UCEC-US: 1/531 | Somatic variant |
| ICGC TCGA | Bladder | chr7:g.17382605C>T | Q822* | BLCA-US: 1/411 | Somatic variant |
| ICGC TCGA COSMIC | Lung | chr7:g.17379124G>T | E559* | LUAD-US: 1/516 | Somatic variant |
| ICGC COSMIC | Esophagus | chr7:g.17379784C>T | Q779* | ESCA-CN: 1/332 | Somatic variant |
| ICGC COSMIC | Ovary | chr7:g.17375370G>T | G374* | OV-US: 1/426 | Somatic variant |
| ICGC COSMIC | Gall Bladder Colorectal | chr7:g.17374617C>T | R339* | BTCA-JP: 1/239 | Somatic variant |
| ICGC TCGA COSMIC | Lung | chr7:g.17378833C>T | Q462* | LUAD-US: 1/516 | Somatic variant |
| ICGC TCGA | Bladder | chr7:g.17378911G>T | E488* | BLCA-US: 1/411 | Somatic variant |
| TCGA | Colorectal | chr7:g.17339506delT | F562Sfs*17 | TCGA-COAD: 1/400 | Somatic variant |
| TCGA | Colorectal | chr7:g.17339471delATTT | D549Vfs*29 | TCGA-COAD: 1/400 | Somatic variant |
| TCGA | Lung | chr7:g.17340116delGTTATGCT... | C764* | TCGA-LUAD: 1/567 | Somatic variant |
| TCGA | Stomach | chr7:g.17339119delA | N433Mfs*15 | TCGA-STAD: 1/440 | Somatic variant |
| TCGA | Stomach | chr7:g.17340012delC | T731Lfs*6 | TCGA-STAD: 1/440 | Somatic variant |

| | | | | | |
|---------------|------------------|--|------------|-------------------------------------|------------------|
| TCGA | Ovary Brain | chr7:g.17339390delC | S522Yfs*22 | TCGA-OV: 1/436 TCGA-GBM: 1/393 | Somatic variant |
| TCGA | Cervix uteri | chr7:g.17330760delCA | T194Wfs*11 | TCGA-CESC: 1/289 | Somatic variant |
| TCGA | Kidney | chr7:g.17310070delT | S68Qfs*10 | TCGA-KIRC: 1/336 | Somatic variant |
| TCGA | Stomach | chr7:g.17310041_17310042insT | Q58Sfs*5 | TCGA-STAD: 1/440 | Somatic variant |
| TCGA | Ovary | chr7:g.17339363_17339364insTTTAC GAA... | Q513Hfs*9 | TCGA-OV: 1/436 | Somatic variant |
| TCGA | Ovary | chr7:g.17334057_17334058insAATGA TCC... | N284Kfs*16 | TCGA-OV: 1/436 | Somatic variant |
| TCGA | Cervix uteri | chr7:g.17339689_17339690insA | Q623Tfs*26 | TCGA-CESC: 1/289 | Somatic variant |
| TCGA | Kidney | chr7:g.17339823_17339824insC | Q667Pfs*7 | TCGA-KIRC: 1/336 | Somatic variant |
| TCGA | Head and neck | chr7:g.17322597_17322598insTGAAA TAC... | L118Efs*11 | TCGA-HNSC: 1/336 | Somatic variant |
| TCGA | Brain | chr7:g.17330075G>T | E192* | TCGA-GBM: 1/393 | Somatic variant |
| COSMIC | Blood | chr7:g.17339252C>G | S461* | NS | Somatic variant |
| COSMIC | lung | chr7:g.17334937C>G | S305* | NS | Unknown origin |
| COSMIC | oesophagus | chr7:g.17339886C>A | Y672* | NS | Somatic variant |
| COSMIC | Brain | chr7:g.17338985G>A | p.? | NS | Somatic variant |
| COSMIC | Colorectal | chr7:g.17339282del | F471Sfs*5 | NS | Somatic variant |
| COSMIC | ovary | chr7:g.17339505dup | F546lfs*5 | NS | Somatic variant |
| COSMIC | stomach | chr7:g.17330773del | Q183Rfs*3 | NS | Unknown origin |
| COSMIC | Colorectal | chr7:g.17339761C>T | Q631* | NS | Unknown origin |
| COSMIC | Colorectal | chr7:g.17339471_17339474del | D534Vfs*28 | NS | Somatic variant |
| COSMIC | skin | chr7:g.17330755G>A | p.? | NS | Somatic variant |
| COSMIC | Gall bladder | chr7:g.17339596A>T | K576* | NS | Somatic variant |
| COSMIC | Colorectal | chr7:g.17339510del | F547Sfs*16 | NS | Somatic variant |
| COSMIC | lung | chr7:g.17339017C>T | R383* | NS | Somatic variant |
| COSMIC | Colorectal | chr7:g.17335773C>T | Q368* | NS | Unknown origin |
| COSMIC | Blood | chr7:g.17343043del | T828Hfs*60 | NS | Somatic variant |
| COSMIC | skin | chr7:g.17339371C>T | Q501* | NS | Somatic variant |
| COSMIC | Brain | chr7:g.17333936A>T | K229* | NS | Somatic variant |
| COSMIC | stomach | chr7:g.17330773dup | Q183Pfs*7 | NS | Somatic variant |
| COSMIC | Blood | chr7:g.17339177_17339178del | L436Qfs*3 | NS | Somatic variant |
| COSMIC | NS | chr7:g.17310124G>A | p.? | NS | Somatic variant |
| COSMIC | kidney | chr7:g.17309946dup | I11Nfs*4 | NS | Somatic variant |
| GnomAD Cancer | NS | chr7:g.17338921del | K14Cfs*5 | GnomAD Exomes: 1/122002 | Germline variant |

Classification of the severity of the variant consequence, based on agreement with VEP. All *AHR* variants assumed to have High impact in the protein, probably causing protein truncation or loss of function. The nomenclature used for the variants annotation is as follow: ICGC and GnomAD “Cancer” (ENST00000242057) |TCGA and COSMIC (ENST00000642825). ESAD-UK: Esophagus, United

Kingdom; OV-US: Ovary, United States; KIRC-US: Kidney, United States; SKCM-US: Skin, United States; STAD-US: Stomach, United States; DLBC-US: Blood, United States; MALY-DE: Blood, Germany; BLCA-CN: Bladder, China; PEME-CA: Brain, Canada; BLCA-US: Bladder, United States; LICA-CN: Liver, China; GBM-US: Brain, United States; COAD-US: Colorectal, United States; BRCA-UK: Breast, United Kingdom; UCEC-US: Uterus, United States; LUAD-US: Lung, United States; ESCA-CN: Esophagus, China; BTCA-JP: Gall Bladder, Japan; TCGA-COAD: Colorectal; TCGA-LUAD: Lung; TCGA-STAD: Stomach; TCGA-OV: Ovary; TCGA-CESC: Cervix uteri; TCGA-KIRC: Kidney; TCGA-HNSC: Head and neck; TCGA-GBM: Brain; NS: not specified. Data obtained after removing duplicates.

Variants in AHR per population

The *AHR* variants were reported across 14 countries by ICGC, 12 race/ethnicities by TCGA and 7 populations by GnomAD. The country, race/ethnicity and population with the highest number of variants were: United States (ICGC: 246 variants – 21.8%), White/Not Hispanic or Latino (TCGA: 124 variants – 54.1%) and European Non Finnish (GnomAD Cancer: 23 variants – 79.3%; GnomAD Non Cancer: 2,023 variants – 46.4%). Also, variants in Black/African American|Not hispanic or Latino and African population represented a great number in TCGA (25 variants – 11%) and GnomAD (GnomAD Cancer: 3 variants – 10.3%; GnomAD Non Cancer: 1,848 variants – 42.6%), respectively (Figure 6).

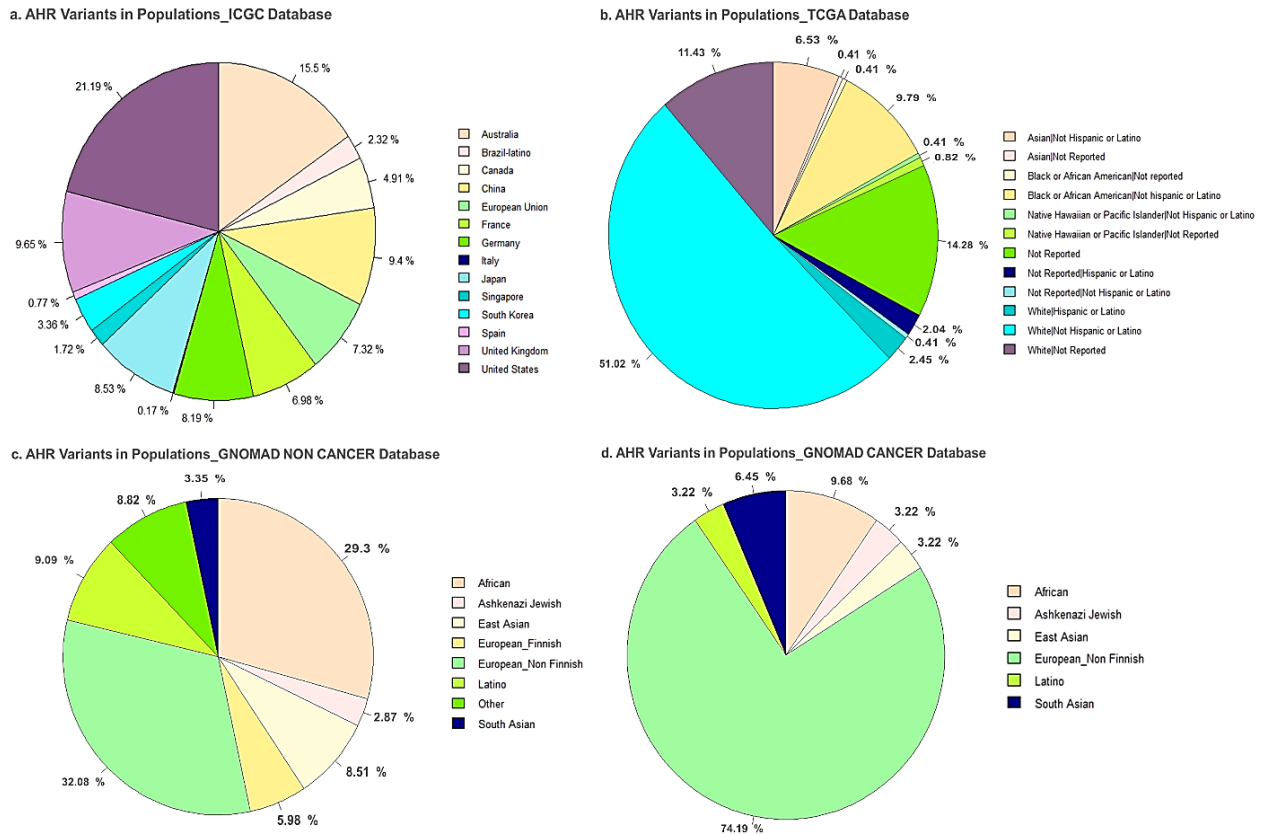


Figure 6. Distribution of variants in the *AHR* gene by population/country. Percentage of variants found in the whole dataset from the genomic repositories after filtering and removing duplicates. Populations were classified as it was reported by each database: a. ICGC (Country), b. TCGA (Race |ethnicity), c. GnomAD Non-Cancer, d. GnomAD “Cancer” (population).

Variants in Hispanic/Latinos

From the 2,079 variants, 39 (1.88%) were present in Hispanic/Latinos. In this population, the primary sites with the highest number of variants in *AHR* were Skin (27 variants – 69.2%) and Uterus (with 5 variants – 12.8%). The most frequent types of variants based on their consequence were intronic (36%) and missense (18%). Evaluation of the variant effects with VEP provided functional impact predictions as follows: modifier (64%), low (8%), moderate (18%) and high impact (10%) (Table 2).

The highest number of cancer variants reported for this population were found in ICGC (27 variants – 2.4%) and TCGA (11 variants – 4.8%).

Table 2. Variants in *AHR* in individuals with Hispanic ancestry. Variants in Hispanic/Latino were reported in ICGC, TCGA, and GNOMAD “Cancer”.

| VEP IMPACT VARIANT CONSEQUENCE | FATHMM IMPACT SCORE | GENOMIC DNA CHANGE | CONSEQUENCE | SOURCE | PRIMARY SITE | SOMATIC OR GERMLINE STATUS |
|--------------------------------------|------------------------------------|---|-----------------------|------------------|--------------------------------|----------------------------------|
| High | Pathogenic (0.98533) | chr7:g.17378635C>T | Stop gained | ICGC TCGA COSMIC | Stomach | Somatic variant |
| High | (-) | chr7:g.17330760delCA | Frameshift variant | TCGA | Cervix | Somatic variant |
| High | (-) | chr7:g.17339363 17339364insTTTACGAA... | Frameshift variant | TCGA | Ovary | Somatic variant |
| High | (-) | chr7:g.17339689 17339690insA | Frameshift variant | TCGA | Cervix | Somatic variant |
| Moderate | Benign (0.138352) | chr7:g.17378911G>A | Missense variant | ICGC TCGA COSMIC | Uterus Colorectal Kidney | Somatic variant |
| Moderate | Pathogenic (0.89964) | chr7:g.17375410C>T | Missense variant | ICGC TCGA COSMIC | Uterus | Somatic variant |
| Moderate | Pathogenic (0.98602) | chr7:g.17379548T>G | Missense variant | ICGC TCGA COSMIC | Uterus | Somatic variant |
| Moderate | Benign (0.080740) | chr7:g.17379146T>C | Missense variant | ICGC TCGA COSMIC | Bladder | Somatic variant |
| Moderate | Benign (0.189142) | chr7:g.17378702A>C | Missense variant | ICGC | Skin | Somatic variant |
| Moderate | Benign (0.158648) | chr7:g.17309953C>T | Missense variant | TCGA | Uterus | Somatic variant |
| Moderate | Pathogenic (0.777609) | chr7:g.17349721A>G | Missense variant | gnomAD Cancer | Unknown | Germline variant |
| Low | Benign_high conf. (0.005716) | chr7:g.17378709T>C | synonymous variant | ICGC COSMIC | Skin Oesophagus | Somatic variant |
| Low | Benign_high conf. (0.009427) | chr7:g.17378775C>T | Synonymous variant | ICGC | Skin | Somatic variant |
| Low | Benign_high conf. (0.005544) | chr7:g.17362177A>G | Synonymous variant | ICGC TCGA COSMIC | Liver | Somatic variant |
| Modifier | (-) | chr7:g.17384936 17384937insCA | 3' UTR variant | ICGC | Skin | Somatic variant |
| Modifier | (-) | chr7:g.17333628 17333629insGTGTATATATA | Upstream variant | ICGC | Skin | Somatic variant |
| Modifier | Benign_high conf. (0.027095) | chr7:g.17355504C>T | Intron variant | ICGC COSMIC | Skin Prostate | Somatic variant |
| Modifier | Benign (0.076545) | chr7:g.17386963C>T | Downstream variant | ICGC | Skin | Somatic variant |
| Modifier | Benign_high conf. (0.020072) | chr7:g.17335100C>T | Upstream variant | ICGC | Skin | Somatic variant |
| Modifier | Benign (0.078195) | chr7:g.17349913A>G | Intron variant | ICGC | Skin | Somatic variant |

| | | | | | | |
|----------|------------------------------------|--------------------|-----------------------|-------------|--------|-----------------|
| Modifier | Benign (0.033750) | chr7:g.17341796A>G | Intron variant | ICGC | Skin | Somatic variant |
| Modifier | Benign_high conf. (0.024878) | chr7:g.17335008G>A | Upstream variant | ICGC | Skin | Somatic variant |
| Modifier | Benign_high conf. (0.018374) | chr7:g.17343044C>T | Intron variant | ICGC | Skin | Somatic variant |
| Modifier | Benign (0.030214) | chr7:g.17346291A>G | Intron variant | ICGC | Skin | Somatic variant |
| Modifier | Benign (0.316295) | chr7:g.17383097T>G | 3' UTR variant | ICGC | Skin | Somatic variant |
| Modifier | Benign_high conf. (0.019941) | chr7:g.17335927G>A | Upstream variant | ICGC | Skin | Somatic variant |
| Modifier | Benign_high conf. (0.026217) | chr7:g.17338791A>C | 5' UTR variant | ICGC | Skin | Somatic variant |
| Modifier | Benign_high conf. (0.024781) | chr7:g.17351919G>A | Intron variant | ICGC COSMIC | Skin | Somatic variant |
| Modifier | Benign (0.032964) | chr7:g.17351558C>T | Intron variant | ICGC | Skin | Somatic variant |
| Modifier | Benign (0.464348) | chr7:g.17383211C>A | 3' UTR variant | ICGC TCGA | Uterus | Somatic variant |
| Modifier | Benign (0.038853) | chr7:g.17380999C>T | Intron variant | ICGC | Skin | Somatic variant |
| Modifier | Benign_high conf. (0.022740) | chr7:g.17342764G>C | Intron variant | ICGC | Skin | Somatic variant |
| Modifier | Benign_high conf. (0.024030) | chr7:g.17364148C>T | Intron variant | ICGC | Skin | Somatic variant |
| Modifier | Benign_high conf. (0.023056) | chr7:g.17384767T>G | 3' UTR variant | ICGC | Skin | Somatic variant |
| Modifier | Benign_high conf. (0.024928) | chr7:g.17375868C>T | Intron variant | ICGC | Skin | Somatic variant |
| Modifier | Benign (0.043372) | chr7:g.17339301G>A | intron variant | ICGC | Skin | Somatic variant |
| Modifier | Benign (0.090820) | chr7:g.17387263C>T | Downstream variant | ICGC | Skin | Somatic variant |
| Modifier | Benign_high conf. (0.017694) | chr7:g.17368859G>A | Intron variant | ICGC | Skin | Somatic variant |
| Modifier | Benign (0.036315) | chr7:g.17359381T>C | Intron variant | ICGC | Skin | Somatic variant |

Total number of variants identified in Hispanic/Latino Population according to classification of the severity of the variant consequence, based on VEP and FATHMM. VEP: high impact variant consequence assumed to have a disruptive impact in the protein; moderate impact variant consequence corresponds to a non-disruptive variant that might change protein effectiveness; low impact variant consequence relates to a variant that is assumed to be mostly harmless or unlikely to change protein behaviour; modifier impact variant consequence usually non-coding variants where predictions are difficult or there is no evidence of impact (<https://www.ensembl.org/Help/Glossary>).

DISCUSSION

Recently, there has been renewed interest in the regulatory roles of the AhR in the immune system and cancer. Studies have shown that AhR expression is upregulated in most tumor cells and is closely associated with tumor proliferation, invasion, metastasis, and immune escape (Stockinger et al. 2014; Xue, Fu & Zhou, 2018; Dai et al, 2022; Elson & Kolluri, 2023). Interestingly, we identified a higher number of variants in *AHR* in different cancer types such as Liver (15.1%), Breast (11%), Skin (10.7%), Esophagus (8.6%), Pancreas (7.5%), Brain (5.9%), Colorectal (5.3%) and Bladder (4.5%), and a total of 41 high impact variants (mostly missense and stop-gained consequences) occurred in these primary sites covering the genomic region around exon 7–9 in AhR (*Chr7:17372389–17378532*). At the protein level, loss of exons 8 and 9 results in an aberrant protein coding sequence, with the probably loss of amino acids, disrupting the PAS-B domain and the C-terminal part of the ligand-binding domain of the protein, as well as the HSP90 and XAP2 protein binding domains which are predicted to be affected too (Vlaar et al, 2022), leading to the activation of AhR pathway independently of ligand binding, which is in line with the increased expression of genes targeted by the receptor.

Also, we found a non-synonymous variant (AhR.Q383H) in *AHR* present in 3 different types of cancer (Liver, Bladder and Esophagus). According to the literature, this variant has been identified as an APOBEC-associated hotspot mutation in

cancer (Shi et al, 2020) which affects the ligand binding affinity and specificity, and together with alteration of exon 8 and 9 lead to a constitutively active AhR signaling inducing an oncogenic phenotype in the cells, which suggest that aberrant AhR signaling may be an important driver of tumorigenesis (Wang et al, 2020; Vlaar et al, 2022) and support the fact that Ah Receptor is chronically activated in many tumor types.

The AhR redox activity has been mainly associated with the detoxification of xenobiotics and pollutants (Denison and Nagy, 2003; Johnson et al. 2020) given the important role of AhR regulating the expression of a large superfamily of antioxidant molecules known as cytochrome p450 proteins (Murray, Patterson and Perdew, 2014). Exposure to xenobiotics is a significant cause of cancer incidence worldwide (Tamási et al, 2011; Wu et al, 2018) and one of the cancer types that has generally been correlated with exposure to tobacco-derived carcinogens, excessive alcohol consumption, or both, is Head and Neck Cancer (HNC) (Denison and Nagy, 2003; Johnson et al. 2020). In particular, Tobacco and tobacco smoke, are rich in polycyclic aromatic hydrocarbons and nitrosamines, known as human carcinogens related to strongly increased risk of HNC (Johnson et al. 2020). These carcinogens undergo metabolic activation, forming reactive metabolites that can cause damage in the DNA by creating bulky DNA adducts. Failure in the processes of detoxification and excretion of these metabolites may result in the permanent damage of DNA leading to the progression of HNC (Johnson et al. 2020). In this context, we identified 19 variants in *AHR* reported in HNC, but none of them were reported in Latin American population. From these variants, seven may have high impact consequences, and from those, 6 variants (86%) occurred in the dimerization domain (bHLH-PAS A

domain), probably affecting the molecular signaling of biotransformation of hydrocarbons by CYP enzymes, targeted by the AhR. The aryl hydrocarbon receptor (AhR) is a transcription factor that can be activated by xenobiotic ligands. Once activated, the AhR forms a heterodimer with the aryl hydrocarbon receptor nuclear translocator (ARNT) and binds to the xenobiotic response element (XRE) to regulate the expression of downstream genes, including cytochrome P450 family 1 subfamily A member 1 (CYP1A1) enzymes that are involved in the correct metabolism of xenobiotics.

Here, we identified 2,079 variants in different tumor types reported in three cancer genomics data repositories; from this number, 64 (3.1%) variants were predicted to have a high impact on the protein function according to the VEP score. The 64 variants included truncating, INDEL and start lost mutations (Table 1). The INDELs located in the bHLH domain, and the frameshift variants and truncating alterations occurring in the PAS-A domain might weaken the formation of a stable AHR-ARNT heterodimer, which is critical for the activation of the AhR signaling pathway, since it recognizes the xenobiotic response elements (XRE) in the promoter of downstream genes. Also, the heterodimer AHR-ARNT is important for the ligand-dependent transformation of the AhR and the displacement of heat shock protein 90 (Hsp90) in the PAS A / B domains (Wu et al. 2013; Schulte et al. 2017). This could be possibly due to interaction interfaces in the bHLH as well as the PAS-A domain, where the amino acids L50, A79, F82, L122, and the ones found in the present study, L118 and A121 (rs746848937) have been identified as crucial for the role of the AHR-ARNT heterodimer for binding to the promoter region of AhR target genes and the

xenobiotic response elements (XRE), and, probably leading to a loss of transcriptional activity of the AhR (Swanson, Chan & Bradfield, 1995; Haidar et al. 2021). Furthermore, most high impact variants in the ligand binding domain - PAS B corresponded to stop gained consequence (p.S305*; p.R339*; p.R359*; p.Q368*; p.G374*; p.R383*), which results in the production of a truncated protein. The PAS-B domain not only acts as a sensor of environmental and physiological signals, it also provides a binding interface for HSP90, XAP2 and the PAS-B domain of ARNT (Dai et al. 2022). However, the AHR-ARNT can stably dimerize without the PAS B domain, but truncation or deletion of the PAS-B is predicted to affect the HSP90 and XAP2 protein binding domains and leads to constitutively activated AhR signaling inducing oncogenic processes in the cell (Vlaar et al. 2022). Additionally, 24 (37.5%) high impact alterations were found in the Transactivation Domain (TAD), corresponding to 13 frameshift and 11 truncating variants, which could likely inactivate this region of the protein. According to the literature, the TAD is required for full AhR function suggesting that the Q-rich/PST domain in the C-terminus is a crucial component for regulating intracellular trafficking and nuclear accumulation after ligand binding (Tkachenko et al. 2016; Haidar et al. 2021).

Thirty-nine *AHR* variants were identified in Latin American populations, 1 (2.6%) in GnomAD Cancer; 24 (61.5%) in ICGC; 4 (10.2%) in TCGA; 1 (2.6%) in ICGC - TCGA; 3 (7.7%) in ICGC - COSMIC; 6 (15.4%) in ICGC - TCGA - COSMIC. Seven genomic alterations were detected as high impact consequence (Table 2). According to the distribution of these high impact variants in the protein two are located in the dimerization domain (bHLH – PAS A: p.Y76C; T194Wfs*11), two in the ligand

binding domain (PAS B: p.T387I; p.R396*) and three in the transcriptional activation domain (TAD: p.Q513Hfs*9; p.Q623Tfs*26; p.F700C), thus high impact consequences may disrupt interaction of the bHLH – PAS domains leading to changes in the function of the protein by producing a constitutively active receptor more potent than the intact protein, which will promotes carcinogenesis (McGuire et al. 2001; Nan et al. 2011).

According to our findings in COSMIC, ICGC and TCGA databases, variants in *AHR* were reported in 24 cancer types, while in Latin American population *AHR* variations were found in Skin cancer and Uterus. Other malignancies have been studied in this population (Colorectal, Kidney, Esophagus, Prostate, Stomach, Liver, Bladder and Ovary), but only one *AHR* variant for each cancer have been informed according to our search. In this context, it is important to highlight, that Latin American populations have been underrepresented in most worldwide cancer studies (Popejoy & Fullerton, 2016; Landy et al. 2018). As shown in Dataset 2, Brazil was the only Latin American cohort in ICGC available for our analysis, and TCGA are biased toward the inclusion of White|Not Hispanic or Latino individuals, and other ethnicities are underrepresented.

The variant chr7: g.17375410C>T (rs1246045330) was the most frequent high impact variant in the Latin American population (Table 2). It was reported by the three cancer genomic repositories (COSMIC, ICGC, TCGA) and the variant was found in two patients with Uterus cancer (2/531 donors affected). This variant has been reported as uncertain significance in ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/RCV002726551/>). No publications for

rs1246045330 are available, and it was not possible to identify if an individual could be found in more than one repository.

Finally, due to the differences in the genetic composition of different population, it is important to perform screening of *AHR* genes variants in the Latin American populations, since this population may hold specific variants that can increase cancer susceptibility (Florio et al. 2020; Petrick et al. 2020; Joko-Fru et al. 2020; Sung et al. 2021). Plus, the incidence rates of several primary cancer sites in transitioning countries in South America, Africa and Asia has increased during the recent years (Florio et al. 2020; Petrick et al. 2020; Joko-Fru et al. 2020; Sung et al. 2021).

CONCLUSIONS

This study compiles genetic variants in *AHR* that have a high impact on several cancer sites reported in White, European, Hispanic, and other populations. ICGC, TCGA, and COSMIC repositories have informed variants in *AHR* for 25 primary sites. However, studies reporting *AHR* variation in the Latin American population are limited to 11 cancer sites: Skin, Uterus, Colorectal, Kidney, Esophagus, Prostate, Stomach, Liver, Bladder, Cervix, and Ovary. The under representation of Latin Americans in cancer and population genomics databases, might be related to the small number of variants that we identified for this population (39 out of 2080 variants), which indicates a necessity of screening *AHR* mutations in Hispanic population.

It is possible that individuals from different populations, including admixed populations, may possess genetic variation promoting genomic changes in the AhR protein function or disruption of the mRNA processing. By identifying genetic risk variants, we can gain a better understanding of the role of *AHR* in cancer and develop more effective prevention and treatment strategies in admixed populations.

Recurrent alterations in the *AHR* gene that impact its DNA binding, dimerization or ligand binding domains can result in the constant activation of AhR signaling. This, in turn, can lead to elevated oxidative metabolism and the formation of reactive oxygen species, ultimately inducing an oncogenic phenotype in cells. Identifying high impact variants in this gene could pave the way for targeted therapies for patients with AHR activating mutations. However, additional research is necessary to fully comprehend the involvement of *AHR* in cancer.

This study has limitations, information about race or ethnicity was not available for all repositories searched and consequently some Latinos may be hidden in those studies. No racial/ethnic data for the samples in COSMIC database was accessible.

Ethical statement

The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committees of Universidad Cooperativa de Colombia (INV2085), and Universidad Industrial de Santander (4110).

Acknowledgments

The study was supported by grant INV2085 from CONADI-Universidad Cooperativa de Colombia and it was received by CCL. This work was carried out under research program of M.S. in basic biomedical sciences from Universidad Industrial de Santander.

Author's Contribution

The authors confirm contribution to the paper as follows: study conception and design: CCL, GOS, VCC, TPN; data collection: TPN and GOS; analysis and interpretation of results: CCL, GOS, VCC, TPN; draft manuscript preparation: CCL, GOS, VCC, TPN. All authors were involved in the revision of the draft manuscript and have agreed to the final content.

Competing interests

No competing interests were disclosed.

REFERENCES

- Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., ... & Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nature methods*, 7(4), 248-249. doi: 10.1038/nmeth0410-248
- Adzhubei, I., Jordan, D. M., & Sunyaev, S. R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Current protocols in human genetics*, 76(1), 7-20. doi: 10.1002/0471142905.hg0720s76
- Alsubait, A., Aldossary, W., Rashid, M., Algamdi, A., & Alrfaei, B. M. (2020). CYP1B1 gene: implications in glaucoma and cancer. *Journal of Cancer*, 11(16), 4652-4661. doi: 10.7150/jca.42669
- Baker, J., Sakoff, J., & McCluskey, A. (2019). The aryl hydrocarbon receptor (AhR) as a breast cancer drug target. *Medicinal Research Reviews*, 40(2), 556-590. doi: 10.1002/med.21645
- Bradshaw, T. D., & Bell, D. R. (2009). Relevance of the Aryl hydrocarbon Receptor (AhR) for clinical toxicology. *Clinical Toxicology*, 47(7), 632-642. doi: 10.1080/15563650903086436
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 68(6), 394-424. doi: 10.3322/caac.21492
- Brunotto, M., Zarate, A. M., Bono, A., Barra, J. L., & Berra, S. (2014). Risk genes in head and neck cancer: a systematic review and meta-analysis of last 5 years. *Oral Oncology*, 50(3), 178-188. doi: 10.1016/j.oraloncology.2013.11.004
- Carrio-Cordo, P., Acheson, E., Huang, Q., & Baudis, M. (2020). Geographic assessment of cancer genome profiling studies. *Database : the Journal of Biological Databases and Curation*, 2020, baaa009. doi: 10.1093/database/baaa009
- Chen, D., Tian, T., Wang, H., Liu, H., Hu, Z., Wang, Y., ... & Lu, D. (2009). Association of human aryl hydrocarbon receptor gene polymorphisms with

risk of lung cancer among cigarette smokers in a Chinese population. *Pharmacogenetics and Genomics*, 19(1), 25-34. doi: 10.1097/FPC.0b013e328316d8d8

Cheng, T., Gamage, S. M. K., Hewage, D., Lu, C. T., Aktar, S., Gopalan, V., & Lam, A. K. Y. (2022). AHR gene expression and the polymorphism rs2066853 are associated with clinicopathological parameters in colorectal carcinoma. *Human Pathology*, 122, 50-59. doi: 10.1016/j.humpath.2021.12.012

Chong, Z. X., Yong, C. Y., Ong, A. H. K., Yeap, S. K., & Ho, W. Y. (2023). Deciphering the Roles of Aryl Hydrocarbon Receptor (AHR) in Regulating Carcinogenesis. *Toxicology*, 153596. doi: 10.1016/j.tox.2023.153596

Dai, S., Qu, L., Li, J., et al. (2022). Structural insight into the ligand binding mechanism of aryl hydrocarbon receptor. *Nature Communications*, 13, 6234. doi: 10.1038/s41467-022-33858-w

Denison, M. S., & Nagy, S. R. (2003). Activation of the aryl hydrocarbon receptor by structurally diverse exogenous and endogenous chemicals. *Annual Review of Pharmacology and Toxicology*, 43, 309-334. doi: 10.1146/annurev.pharmtox.43.100901.135828

Feng, S., Cao, Z., & Wang, X. (2013). Role of aryl hydrocarbon receptor in cancer. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 1836(2), 197-210. doi: 10.1016/j.bbcan.2013.05.001

Fernández-Mateos, J., Seijas-Tamayo, R., Adansa Klain, J. C., Pastor Borgoñón, M., Pérez-Ruiz, E., Mesía, R., ... & Cruz-Hernández, J. J. (2019). Genetic susceptibility in head and neck squamous cell carcinoma in a Spanish population. *Cancers*, 11(4), 493. doi: 10.3390/cancers11040493

Forbes, S.A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., et al. (2015). COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Research*, 43(D1), D805-D811. doi: 10.1093/nar/gku1075

Gagliani, N., Vesely, M.C.A., Iseppon, A., Brockmann, L., Xu, H., Palm, N.W., ... & Akbari, O. (2015). TH17 cells transdifferentiate into regulatory T cells during resolution of inflammation. *Nature*, 523(7559), 221-225. doi: 10.1038/nature14452

- Galle-Treger, L., Suzuki, Y., Patel, N., Sankaranarayanan, I., Aron, J.L., Maazi, H., ... & Akbari, O. (2016). Nicotinic acetylcholine receptor agonist attenuates ILC2-dependent airway hyperreactivity. *Nature Communications*, 7(1), 13202. doi: 10.1038/ncomms13202
- Gudmundsson, S., Singer-Berk, M., Watts, N.A., Phu, W., Goodrich, J.K., Solomonson, M., ... & O'Donnell-Luria, A. Human Mutation (2021). Variant interpretation using population databases: Lessons from gnomAD. *Human Mutation* 1-19 (2021). doi:10.1002/humu.24309
- Haidar, R., Henkler, F., Kugler, J., et al. (2021). The role of DNA-binding and ARNT dimerization on the nucleo-cytoplasmic translocation of the aryl hydrocarbon receptor. *Scientific Reports*, 11, 18194. doi: 10.1038/s41598-021-97507-w
- Hanieh, H., Mohafez, O., Hairul-Islam, I., Alzahrani, A., Bani Ismail, M., & Thirugnanasambantham, K. (2016). Novel Aryl hydrocarbon Receptor agonist suppresses migration and invasion of breast cancer cells. *PLoS One*, 11(1), e0146545. doi: 10.1371/journal.pone.0146545
- Heberle, H., Meirelles, G. V., da Silva, F. R., Telles, G. P., & Minghim, R. (2015). InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics*, 16, 169. doi: 10.1186/s12859-015-0611-3
- Hu, J., Ding, Y., Liu, W., & Liu, S. (2023). When AHR signaling pathways meet viral infections. *Cellular Communication and Signaling*, 21(1), 45. doi: 10.1186/s12964-023-01058-8
- Joko-Fru, W. Y., Jedy-Agba, E., Korir, A., et al. (2020). The evolving epidemic of breast cancer in sub-Saharan Africa: results from the African Cancer Registry Network. *International Journal of Cancer*, 147(8), 2131-2141. doi: 10.1002/ijc.32975
- Karman, B. N., Flaws, J. A., & Herna, I. (2009). The role of the Aryl hydrocarbon Receptor in the female reproductive system. *Biochemical Pharmacology*, 77(4), 547-559. doi: 10.1016/j.bcp.2008.10.020
- Kumar, P., Henikoff, S., & Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, 4(7), 1073-1081. doi: 10.1038/nprot.2009.86

- Landry, L. G., Ali, N., Williams, D. R., Rehm, H. L., & Bonham, V. L. (2018). Lack of diversity in genomic databases is a barrier to translating precision medicine research into practice. *Health Affairs*, 37(5), 780-785. doi: 10.1377/hlthaff.2017.1474
- Lafleur, V.N., Halim, S., Choudhry, H., Ratcliffe, P.J., & Mole, D.R. (2023). Multi-level interaction between HIF and AHR transcriptional pathways in kidney carcinoma. *Life Science Alliance*, 6(4), e202201756. doi: 10.26508/lsa.202201756
- McGuire, J., Okamoto, K., Whitelaw, M.L., Tanaka, H., & Poellinger, L. (2001). Definition of a dioxin receptor mutant that is a constitutive activator of transcription: delineation of overlapping repression and ligand binding functions within the PAS domain. *Journal of Biological Chemistry*, 276(46), 41841-41849. doi: 10.1074/jbc.M106799200
- McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R., Thormann, A., ... & Cunningham, F. (2016). The ensembl variant effect predictor. *Genome Biology*, 17(1), 1-14. doi: 10.1186/s13059-016-0974-4
- Miller, P. D., et al. (2002). Combination of the variant genotype of GSTP1, GSTM1 and p53 are associated with an increased lung cancer risk. *Cancer Research*, 62, 2819-2823.
- Mulero-Navarro, S., & Fernandez-Salguero, P. M. (2016). New trends in aryl hydrocarbon receptor biology. *Frontiers in Cell and Developmental Biology*, 4, 45. doi: 10.3389/fcell.2016.00045
- Murray, I. A., Patterson, A. D., & Perdew, G. H. (2014). Aryl hydrocarbon receptor ligands in cancer: Friend and foe. *Nature Reviews Cancer*, 14, 801-814. doi: 10.1038/nrc3846
- Hao, N., et al. (2011). Identification of residues in the N-terminal PAS domains important for dimerization of Arnt and AhR. *Nucleic Acids Research*, 39(9), 3695-3709. doi: 10.1093/nar/gkq1336
- Norppa, H. (2003). Genetic susceptibility, biomarkers response, and cancer. *Mutagenesis Research*, 544, 339-348.
- Popejoy, A. B., & Fullerton, S. M. (2016). Genomics is failing on diversity. *Nature*, 538(7624), 161-164. doi: 10.1038/538161a

- Popolo, A., Pinto, A., Daglia, M., Nabavi, S. F., Farooqi, A. A., & Rastrelli, L. (2017). Two likely targets for the anti-cancer effect of indole derivatives from cruciferous vegetables: PI3K/Akt/mTOR signalling pathway and the aryl hydrocarbon receptor. *Seminars in Cancer Biology*, 46, 132-137. doi: 10.1016/j.semcancer.2017.05.006
- Powell, J. B., Goode, G. D., & Eltom, S. E. (2013). The Aryl Hydrocarbon Receptor: A Target for Breast Cancer Therapy. *Journal of Cancer Therapy*, 4(7), 1177-1186. doi: 10.4236/jct.2013.47137
- R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing.
- R Studio Team. (2019). RStudio: Integrated Development Environment for R.
- Re, A., Ferraù, F., Cafiero, C., Spagnolo, F., Barresi, V., Romeo, D. P., ... & Cannavò, S. (2020). Somatic Deletion in Exon 10 of Aryl Hydrocarbon Receptor Gene in Human GH-Secreting Pituitary Tumors. *Frontiers in Endocrinology*, 11, 591039. doi: 10.3389/fendo.2020.591039
- Reva, B., Antipin, Y., & Sander, C. (2007). Determinants of protein function revealed by combinatorial entropy optimization. *Genome Biology*, 8(11), R232. doi: 10.1186/gb-2007-8-11-r232
- Reva, B., Antipin, Y., & Sander, C. (2011). Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Research*, 39(17), e118-e118. doi: 10.1093/nar/gkr407
- Roman ÁC, Carvajal-Gonzalez JM, Merino JM, Mulero-Navarro S, Fernández-Salguero PM. (2018). The aryl hydrocarbon receptor in the crossroad of signalling networks with therapeutic value. *Pharmacology & Therapeutics*, 185, 50-63. doi: 10.1016/j.pharmthera.2017.12.007
- Safe, S., Cheng, Y., & Jin, U. H. (2017). The Aryl Hydrocarbon Receptor (AhR) as a Drug Target for Cancer Chemotherapy. *Current Opinion in Toxicology*, 2, 24-29. doi: 10.1016/j.cotox.2017.01.012
- Sato, Y., Fujimura, T., Tanita, K., Chunbing, L., Matsushita, S., Fujisawa, Y., ... & Aiba, S. (2019). Malassezia-derived aryl hydrocarbon receptor ligands enhance the CCL20/Th17/soluble CD163 pathogenic axis in extramammary

- Paget's disease. *Experimental Dermatology*, 28(8), 933-939. doi: 10.1111/exd.13944
- Shi, M.-J., et al. (2020). Identification of new driver and passenger mutations within APOBEC-induced hotspot mutations in bladder cancer. *Genome Medicine*, 12(1), 85.
- Shihab, H. A., Gough, J., Cooper, D. N., Stenson, P. D., Barker, G. L., Edwards, K. J., ... & Gaunt, T. R. (2013). Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Human Mutation*, 34(1), 57-65. doi: 10.1002/humu.22225
- Schulte, K. W., Green, E., Wilz, A., Platten, M., & Daumke, O. (2017). Structural basis for aryl hydrocarbon receptor-mediated gene activation. *Structure*, 25(7), 1025-1033.e3. doi: 10.1016/j.str.2017.05.005
- Sollis, E., Mosaku, A., Abid, A., Buniello, A., Cerezo, M., Gil, L., ... & Harris, L. W. (2023). The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Research*, 51(D1), D977-D985. doi: 10.1093/nar/gkac1010
- Stejskalova, L., & Pavek, P. (2011). The Function of Cytochrome P450 1A1 Enzyme (CYP1A1) and Aryl Hydrocarbon Receptor (AhR) in the Placenta. *Current Pharmaceutical Biotechnology*, 12(5), 715-730. doi: 10.2174/138920111795470994
- Stockinger, B., Di Meglio, P., Gialitakis, M., & Duarte, J. H. (2014). The aryl hydrocarbon receptor: Multitasking in the immune system. *Annual Review of Immunology*, 32, 403–432.
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 71(3), 209-249.
- Swanson, H. I., Chan, W. K., & Bradfield, C. A. (1995). DNA binding specificities and pairing rules of the Ah receptor, ARNT, and SIM proteins. *Journal of Biological Chemistry*, 270(44), 26292–26302.
- Tamási, V., Monostory, K., Prough, R. A., & Falus, A. (2011). Role of xenobiotic metabolism in cancer: Involvement of transcriptional and miRNA regulation of

P450s. *Cell and Molecular Life Sciences*, 68, 1131–1146. doi: 10.1007/s00018-010-0600-7

Tkachenko, A., Niskanen, E. A., Vashchinkina, E., Kazantseva, M., Schulte, T., & Moore, H. M. (2016). The Q-rich/PST domain of the AHR regulates both ligand-induced nuclear transport and nucleocytoplasmic shuttling. *Scientific Reports*, 6, 32009.

Tian Yang, Ya-Long Feng, Lin Chen, Nosratola D. Vaziri & Ying-Yong Zhao. (2019). Dietary natural flavonoids treating cancer by targeting aryl hydrocarbon receptor. *Critical Reviews in Toxicology*. doi: 10.1080/10408444.2019.1635987

Tomczak, K., Czerwińska, P., & Wiznerowicz, M. (2015). The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemporary Oncology (Poznań)*, 19(1A), A68-A77. doi: 10.5114/wo.2014.47136

Tomkiewicz, C., Herry, L., Bui, L. C., Metayer, C., Bourdeloux, M., Barouki, R., & Coumoul, X. (2013). The aryl hydrocarbon receptor regulates focal adhesion sites through a non-genomic FAK/Src pathway. *Oncogene*, 32, 1811–1820.

Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M., & Ng, P. C. (2016). SIFT missense predictions for genomes. *Nature Protocols*, 11(1), 1-9. doi: 10.1038/nprot.2015.123

Vlaar, J. M., Borgman, A., Kalkhoven, E., & van der Zwan, Y. G. (2022). Recurrent exon-deleting activating mutations in AHR act as drivers of urinary tract cancer. *Scientific Reports*, 12, 10081. doi: 10.1038/s41598-022-14256-0

Wang, Z., Li, M., Zhou, Y., & Wei, X. (2021). How the AHR became important in cancer: The role of chronically active AHR in cancer aggression. *International Journal of Molecular Sciences*, 22(1), 387. doi: 10.3390/ijms22010387

Wei, T., Liu, L., & Zhou, X. (2018). Cortex dictamni extracts inhibit over-proliferation and migration of rat airway smooth muscle cells via FAK/p38/ Bcl-2 signaling pathway. *Biomedicine & Pharmacotherapy*, 102, 1-8. doi: 10.1016/j.biopha.2018.03.047

Wilson, S., Jones, L., & Coussens, C. (2002). The Links Between Environmental Factors, Genetics, and the Development of Cancer. In *Institute of Medicine (US) Roundtable on Environmental Health Sciences, Research, and Medicine*

(Eds.), *Cancer and the Environment: Gene-Environment Interaction* (pp. 43-58). Washington (DC): National Academies Press (US). doi: 10.17226/10149

Wu, D., Potluri, N., Kim, Y., & Rastinejad, F. (2013). Structure and dimerization properties of the aryl hydrocarbon receptor PAS-A domain. *Molecular and Cellular Biology*, 33(21), 4346-4356. doi: 10.1128/MCB.00844-13

Wu, S., Zhu, W., Thompson, P., & Hannun, Y. A. (2018). Evaluating intrinsic and non-intrinsic cancer risk factors. *Nature Communications*, 9(1), 3490. doi: 10.1038/s41467-018-05467-z

Xue, P., Fu, J., & Zhou, Y. (2018). The Aryl Hydrocarbon Receptor and Tumor Immunity. *Frontiers in Immunology*, 9, 286. doi: 10.3389/fimmu.2018.00286

Zhang, J., Bajari, R., Andric, D., et al. (2019). The International Cancer Genome Consortium Data Portal. *Nature Biotechnology*, 37(4), 367-369. doi: 10.1038/s41587-019-0055-9