Unraveling the genetic architecture of hepatoblastoma risk: birth defects and increased burden of germline damaging variants in gastrointestinal/renal cancer predisposition and DNA repair genes

Talita Aguiar*[1], Anne Caroline Teixeira[2], Marília Scliar[3], Juliana Sobral[4], Renan B. Lemes[1], Silvia Souza Costa[2], Giovanna Tolezano[5], Fernanda Francisco[6], Israel Tojal[6], Mônica Cypriano[7], Silvia Toledo[8], Eugênia Valadares[9], Raquel Pinto[10], Osvaldo Artigalás[11], Joaquim Caetano de Aguirre Neto[12], Estela Novak[13], Lilian Maria Cristofani[14], Sofia Sugayama[15], Vicente Odone[16], Isabela Werneck[17], Cecilia da Costa[18], Carla Rosenberg[4], and Ana Krepischi[2]

[1]University of Sao Paulo
[2]Universidade de Sao Paulo Instituto de Biociencias
[3]University of São Paulo
[4]Institute of Biosciences, University of Sao Paulo
[5]University of Sao Paulo Institute of Biosciences Department of Genetics and Evolutionary Biology
[6]ACCamargo Cancer Center
[7]GRAACC
[8]Genetics Laboratory, Pediatric Oncology Institute GRAACC/UNIFESP
[9]Benjamim Guimarães Foundation
[10]Hospital Nossa Senhora da Conceição
[11]Grupo Hospitalar Conceição
[12]Santa Casa de Misericórdia de Belo Horizonte
[13]Fundação Pró-Sangue, Hemocentro de São Paulo
[14]Universidade de São Paulo Faculdade de Medicina
[15]Medical School
[16]Sao Paulo University Medical School
[17]Rede D'OR-São Luiz
[18]Fundação Antônio Prudente

November 12, 2021

Abstract

The ultrarare hepatoblastoma is the most common pediatric liver cancer. HB risk is related to a few rare syndromes, and the molecular bases remain elusive for most cases. We investigated the burden of rare damaging germline variants in 30 Brazilian patients with HB. A high frequency of prematurity (20%) and birth defects (37%), especially craniofacial (17%, including craniosynostosis) and kidney (7%) anomalies, was observed. Pathogenic or likely pathogenic variants mapped to 10 cancer predisposition genes (APC, CHEK2, DROSHA, ERCC5, FAH, MSH2, MUTYH, RPS19, TGFBR2 and VHL,) were detected in 33% of the patients, only 40% of them with a family history of cancer. These findings showed a predominance of

CPGs with a known link to gastrointestinal/colorectal and renal cancer risk. A remarkable feature was an enrichment of rare damaging variants affecting different classes of DNA repair genes, particularly those known as Fanconi anemia genes. Moreover, several damaging rare variants mapped to genes impacting liver functions were observed. To our knowledge, this is the first comprehensive assessment of rare germline variants in HB patients, contributing to elucidating the genetic architecture of HB risk.

INTRODUCTION

The etiology of pediatric cancer is largely unknown [1]. Despite intensive research, gaps remain in our understanding of the genetic landscape of pediatric cancer susceptibility. It is assumed that germline mutations in cancer-predisposing genes (CPGs) in children and adolescents are rare events [2–4]. However, the genetic predisposition to childhood cancer is most likely underidentified. Most investigations have focused on known CPGs and sequenced patients without parental samples, an approach that impairs a broad evaluation of the full range of genetic mechanisms underlying pediatric cancer risk, such as *de novo* mutations and the identification of new candidate CPGs. Recent studies of large cohorts of pediatric cancer patients have confirmed that approximately 8-18% of patients carry a germline pathogenic variant in a broad spectrum of known CPGs [3, 5–10]. These studies also highlighted that isolated factors, such as tumor type and a positive family history of cancer, have low predictive power for the presence of germline CPG mutations.

Hepatoblastoma (HB) is the most common malignant liver tumor in the pediatric population [11], although it is considered an ultrarare disease, accounting for only 1% of all pediatric tumors [12–14]. In Brazil, collected data on HB are concordant with the worldwide incidence of 0.5 to 1.5 cases per million [15, 16]. Most cases are diagnosed before the age of 4 years, and a male preponderance is reported [17]. Nongenetic factors known to be associated with HB risk are related to very low birth weight (<1500 g), including preterm birth (<33 weeks), small for gestational age, and multiple birth pregnancies[11, 18]. A slow increase in HB incidence is observed in North America and Europe [19], which can be partly due to the increased survival of children with low birth weight [20]. An increased risk for HB development has been reported in association with a few specific genetic conditions, including Beckwith-Wiedemann syndrome [21, 22], familial adenomatous polyposis (*APC* gene) [23], Li-Fraumeni syndrome (*TP53* gene) [24], Aicardi syndrome [25], and trisomy 18 [26].

Here, we investigated the germline exome of 30 children who developed HB, 13 of whom (43%) exhibited additional clinical signs. Our analysis provides a framework for investigating candidate genes involved in HB predisposition, as well as the tumor association with specific birth defects.

PATIENTS AND METHODOLOGY

Participants

Thirty children diagnosed with hepatoblastoma were enrolled in this study, which was performed at the Institute of Biosciences, University of São Paulo, Brazil. Patients were recruited from five different Brazilian institutions, most of them (n=27) from three large pediatric cancer centers of the city of São Paulo, namely, the A. C. Camargo Cancer Center, Adolescent and Child with Cancer Support Group (GRAACC), and Pediatric Cancer Institute (ITACI). In addition, three patients were recruited from other institutions, including the Hospital da Baleia (n=1), Hospital da Criança Conceição (n=1), and Hospital São Lucas (n=1). This study was approved by the Research Ethics Committee of the Institute of Biosciences (CAAE: 09163818.4.0000.5464). After signed informed consent was obtained from the parents, peripheral blood samples were collected from 28 patients, and normal liver tissues were recovered from two patients. Genomic samples were also collected from 27 available parents (19 patients).

Library preparation and whole-exome sequencing (WES)

DNA samples were extracted by phenol-chloroform followed by ethanol precipitation [27]. Genomic libraries were constructed with 1 µg of genomic DNA using the following kits: Sureselect QXT V6 (Agilent Technologies), OneSeq Constitutional Research Panel (Agilent Technologies), or xGen Exome Research

Panel v1.0 (IDT - Integrated DNA Technologies). The sequencing of enriched libraries was performed on the Illumina HiSeq 2500 platform using 150 base paired-end reads. The sequences were aligned to the GRCh37/hg19 human genome reference with the BWA_MEM algorithm [28]. Picard tools (v.1.8, http://broadinstitute.github.io/picard/) were used to convert the SAM file into BAM and to mark PCR duplicates. The Genome Analysis Toolkit (GATK 3.7) [29] was used to realign indels, recalibrate the bases, and call (Unified Genotyper) and recalibrate variants (VQSR). Finally, multiallelic variants were split into different lines using the script split_multiallelic_rows.rb from Atlas2 [30] to obtain the VCF files used for analysis.

WES data analysis

SNV and indel variant annotation were conducted using VarSeq software version 1.5.0 (Golden Helix) and selected by the reading depth (>10), Phred score (>20), and alternative allele frequency (>0.35). Based on the public variant databases in ABraOM (http://abraom.ib.usp.br - [31], GnomAD (https://gnomad.broadinstitute.org - [32], and 1000 Genomes - Phase 3 (https://www.internationalgenome.org - [33]), we filtered out germline variants with frequencies above 0.5% or 1% (for dominant and recessive models of inheritance, respectively), as well as those mapped to hypervariable genes [34] or detected in an in-house dataset including data from 19 healthy controls.

Coding variants - Coding nonsynonymous missense and loss-of-function (LoF; frameshift, stop loss/gain, essential splice site, nonsense) variants were maintained for further analysis. *In silico* pathogenicity prediction for missense variants was based on six algorithms provided by the database dbNSFP (version 2.4); those predicted to be damaging to protein function by at least five different tools were prioritized. The final set of genes with rare damaging coding variants was annotated using Varelect [35] and HPO [36] for ranking in association with specific phenotypes. All LoF and prioritized missense variants were validated by visual inspection of the BAM files, further annotated using the Varsome tool [37], and classified according to the American College of Medical Genetics and Genomics (ACMG) guidelines [38, 39].

Noncoding variants - Intronic, intergenic, 3' and 5' untranslated regions (UTRs), and splice region variants were annotated using SNPnexus v4 [40], a web-based annotation tool for the analysis and interpretation of variants that includes databases of regulatory elements and regions, such as miRbase (ftp://mirbase.org/pub/mirbase/20/genomes/), Vista HMR (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/), and ENCODE (ftp://ftp.ensembl.org/pub/grch37/release-95/mysql/regulation_mart_95/); phenotype and disease association (Genetic Association of Complex Diseases and Disorders (GAD) - http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/, ClinVar, and COSMIC); and noncoding scoring (Combined Annotation Dependent Depletion (CADD) - http://cadd.gs.washington.edu/download, Fitness Consequences of Functional Annotation (fitCons) - http://compgen.cshl.edu/fitCons/0downloads/tracks/V1.01/i6/scores/; and Chromatin Effects of Sequence Alterations (DeepSEA) - http://deepsea.princeton.edu/help/).

Copy number variants (CNVs) - CNVs and region of homozygosity (ROH) events were derived from WES data using the software Nexus Copy Number 9 (Biodiscovery) with the SNP-FASST2 segmentation algorithm (threshold $\log_2$ Cy3/Cy5 ratio of |0.3| for gains and losses; minimum ROH length of 5 Mb). Common CNVs (Database of Genomic Variants, http://dgv.tcag.ca/dgv/app/home) were disregarded. For CNV validation, chromosome microarray analysis (CMA) was performed using the 180K platform (Agilent Technologies), as previously reported [41].

RESULTS

Clinical characterization of the HB cohort

The clinical features of the 30 patients who developed HB are described in **Table 1** ; the details of their tumors can be found in*Supplementary Table 1* . The mean age at HB diagnosis was 24 months, excluding one

patient who was diagnosed at 17 years (P07). Twenty-one patients were male (70%), which is in agreement with the literature on sex bias in HB [14, 42]. Six patients (˜20%) presented with a family history of cancer (relatives of different degrees developed different tumors at varying ages).

Fifteen patients were diagnosed with high-risk HB, classified according to the CHIC stratification [43, 44], and 10 presented pulmonary metastasis at diagnosis. Except for P23, who underwent surgery at diagnosis, all patients received neoadjuvant chemotherapy protocols followed by tumor resection or transplantation. Seven patients received a liver transplant, and two relapsed. Four patients died from the disease.

Six out of the 30 patients were born prematurely (<37 weeks), corresponding to 20% of the group. Eleven patients had birth defects (37%), and eight of them were classified as syndromic due to the presentation of two or more congenital clinical features. Among them, five patients had craniofacial anomalies, two of whom were diagnosed with craniosynostosis (P13 and P28); two patients were born with kidney anomalies (P11 and P09); and P07, who was diagnosed with HB at an advanced age (17 years), was born with mild hepatomegaly.

P11, female, had a congenital HB diagnosed at one month of age; in addition, this patient was born with unilateral renal agenesis. P13, a male, was born extremely preterm at 27 weeks. More details about the clinical features of both patients can be found in our previous study [15].

Two patients were born with Hirschsprung disease and other clinical features (P17 and P18). P17, female, was the third child of no consanguineous parents. She was born at term, and her two siblings had a normal phenotype. Abdominal volume and no evacuation were detected in the first 24 hours after birth, and the diagnosis of Hirschsprung disease was made. In the clinical evaluation at 5 years old, she presented with global neuropsychomotor delay, facial dysmorphisms, clinodactyly, and nail dysplasia (hypoplastic). Hepatoblastoma was diagnosed at the age of four and classified as an epithelial subtype with a predominance of embryonal cells, PRETEXT IV, high risk. She underwent the SIOPEL 6 chemotherapy protocol and died before the surgical procedure. P18, male, was the third child of a consanguineous couple; his sister was born with congenital bilateral cataracts, while his brother exhibited intestinal atresia-terminal ileus. The patient was born prematurely (28 weeks), with a syndromic phenotype composed of congenital ileal atresia, bilateral cataracts, and sensorineural deafness. His mother, who had gestational risk (cardiac defect and preeclampsia), died during his birth due to congestive heart failure. Hepatoblastoma tumors were diagnosed at one year of age and classified as fetal epithelial subtype, PRETEXT II, and low risk. He underwent a chemotherapy protocol with cisplatin, doxorubicin, and ifosfamide, followed by partial hepatectomy. Currently, the patient is in post treatment follow-up, and clinical details have been previously published [45].

P24 had facial dysmorphisms and dysplastic nails of the hands and feet, in addition to developmental delay. P29 was born with congenital malformations of the VACTERL spectrum and exhibited postnatal microcephaly and developmental delay. P30 presented craniofacial dysmorphisms, turricephaly, a short neck, laryngomalacia, a swallowing disorder, severe bronchodysplasia, and polysyndactyly of the right 5^{th} digit, in addition to severe malnutrition and neuropsychomotor delay.

Germline coding and noncoding variants (SNVs and indels)

Figure 1 summarizes the analysis workflow of the WES data. The mean sequencing depth of the exomes was 92× (*Supplementary Table 2* presents sequencing metrics and the type of genomic library for each sample); P03 was the only sample presenting 10x on-target coverage below 80%.

A total of 9,467 rare (population frequency <1%) germline coding nonsynonymous variants were detected in the cohort of 30 HB patients, mapping to 6,102 genes. Details of all variants can be found in *Supplementary Table 3* . A total of 2,107 of these rare variants, related to 1,737 different genes and including 1,671 missense mutations and 436 LoF variants (**Figure 2a**), met our criteria of a read depth >10, Phred score >20, and alternative allele frequency >0.35. Pathogenic (P) or likely pathogenic (LP) variants mapped to morbid OMIM genes that could explain the syndromic phenotypes of some patients were not detected.

Using a list of 222 CPGs composed of 119 known CPGs (49 of them reported in OMIM), in addition to 103

candidates that were compiled by revision of recent publications (*Supplementary Table 4*), we investigated the presence of rare coding variants that could be related to cancer development. No homozygous or compound heterozygous pathogenic (P) or likely pathogenic (LP) variants were observed in known or candidate CPGs. Eleven heterozygous P/LP variants mapped to 11 CPGs were detected in 10 patients, comprising 33% of the group (**Table 2**);*VHL* variants were detected in two patients. Among these 10 patients with P/LP in CPGs, only four presented a family history of cancer; four of them were syndromic, and three were born prematurely. One of these patients (P28) carried two variants mapped to known recessive CPGs. Eight out of the ten P/LP variants were detected in seven autosomal-dominant CPGs (known or candidate), including an*APC* LoF variant (OMIM #175100 FAMILIAL ADENOMATOUS POLYPOSIS 1; gastrointestinal carcinomas) and six missense variants mapped to the*CHECK2* (LI-FRAUMENI SYNDROME 2; colorectal, breast and prostate cancer)*, DROSHA* (50)*, MSH2* (LYNCH SYNDROME I; colorectal cancer/MISMATCH REPAIR CANCER SYNDROME 2; hematologic malignancy, brain tumors, and gastrointestinal tumors)*, RPS19* (DIAMOND-BLACKFAN ANEMIA 1; osteogenic sarcoma, myelodysplastic syndrome, colon cancer), *VHL* (VON HIPPEL-LINDAU SYNDROME; renal cell carcinoma, pheochromocytoma, hemangioblastoma, hypernephroma, pancreatic cancer, paraganglioma, adenocarcinoma of the ampulla of Vater)*,* and*TGFBR2* (COLORECTAL CANCER, HEREDITARY NONPOLYPOSIS, TYPE 6) genes. Three P/LP variants were detected in three CPGs associated with recessive conditions: an *ERCC5* LoF (XERODERMA PIGMENTOSUM, COMPLEMENTATION GROUP G; skin cancers) and missense variants mapped to*FAH* (TYROSINEMIA, TYPE I; hepatocellular carcinoma) and*MUTYH* (FAMILIAL ADENOMATOUS POLYPOSIS 2; colorectal carcinomas)*.* In addition, 44 VUS mapped to 34 known/candidate CPGs were observed in 21 patients (70%) (**Table 3**), most of whom carried more than one variant. VUS mapped to the *ATM* ,*BRCA2* , *COL7A1* , *DHCR7* , *DOCK8* , *FANC* s, and *GLI3* genes were detected in more than one patient.

Thirty-two genes related to liver differentiation or function were found to be affected by rare damaging variants *(ABCB11, ABCB4, ABCC2, ABCC3, AFP, AHR, ALB, CTNNB1, CYP1A1, CYP1A2, CYP2C19, CYP2C8, CYP2C9, CYP2D6, CYP3A4, CYP3A7, DLK1, FAH, FOXA2, G6PC, HIF1A, KRT7, KRT8, MET, NR1I2, ONECUT1, PAH, POU5F1, PPARg, SOX17, UGT1A6,* and *UGT1A9).*

We also investigated whether the observed sex bias in the group could be explained by an increased burden of rare damaging variants in one of the sexes; we did not detect significant differences considering the average of rare damaging variants in male and female patients (˜66.7 and 68.1, respectively), LoF variants (˜15 variants in both groups), and rare damaging CPG variants (˜8 and 6, respectively).

A total of 2,069 noncoding variants passed our filters (read depth >10, Phred score >20, alternative allele frequency >0.35, frequency in population databases > 0.1%), including intronic (76%), intergenic (7%), 3' prime UTR (7%), 5' prime UTR (6%), and splice region (4%) variants (*Supplementary Table 5* ; **Figure 2b**). These variants were annotated using SNP Nexus [40], and those with CADD scores above 15 and associated with cancer (https://geneticassociationdb.nih.gov/) were prioritized for further analyses. **Table 4** details the 23 prioritized rare noncoding variants. Two variants were observed in the intronic regions of the CPGs *BRAF* and *CREBBP* , but with no evidence of a functional effect. In particular, P12 carried a *de novo* mutation in the 5' UTR of the *TCF7* gene, an important effector protein in the 7Wnt pathway; this patient also carries a paternally inherited coding *TCF7* VUS (c.1060C>G).

Using the exome data of HB patients and a healthy control group (n=19, data not shown), we inspected a list of 220 DNA repair genes (distributed in 16 categories, including several *bona fide* CPGs; https://www.mdanderson.org/documents/Labs/Wood-Laboratory/human-dna-repair-genes.html), searching for rare LoF and missense variants with high *in silico*damage prediction (5 or more algorithms; *Supplementary Table 6*).**Figure 3** shows the frequency of rare damaging variants detected in 12 DNA repair categories in both controls and patients. Thirty-four heterozygous variants mapped to DNA repair genes were observed in 21 patients (70%), while nine heterozygous variants were detected in nine healthy controls (47%). Although not statistically significant, there was an apparent excess of damaging variants mapped to DNA repair genes in patients. Moreover, rare damaging variants affecting specific DNA repair gene categories, such as ubiquitin modification, poly (ADP-ribose) polymerase (PARP) enzymes that bind

to DNA, nonhomologous end-joining, homologous recombination (*BRCA1, EME2, SPIDR,* and*RAD54L*), Fanconi anemia genes (*BRCA2, FANCA, BRIP1, SLX4, FANCD2,* and *FAAP24*), genes associated with DNA sensitivity to damaging agents, and base excision repair, were observed only in patients. Significant enrichment for rare damaging variants mapped to Fanconi anemia genes was detected in the group of patients (p value 0.0338; Fisher's test).

Germline copy number variation (CNV)

Assessing germline CNVs using exome data, seven rare copy number changes were detected, four with clinical relevance, all of which were validated by CMA. The encompassed genomic regions did not include CPGs or genes associated with the patients' phenotypes (**Table 5**). A pathogenic Y chromosome aneuploidy was detected in P19 (Jacob syndrome 47, XYY). P21 and P28 paternally inherited 15q11.2 copy number gain and loss, respectively, which are classified as risk factors for neurodevelopmental disorders, mainly global delay and intellectual disability.

DISCUSSION

We explored the clinical features and spectrum of rare germline variants possibly associated with cancer risk in 30 children who developed hepatoblastoma. We report here that 37% of HB patients presented with birth defects, mainly craniofacial (17%, including craniosynostosis) and kidney (7%) anomalies, as well as Hirschsprung disease (7%) and nail dysplasia (7%). There is a growing robust bulk of evidence emphasizing that various birth defects occur in association with a significant increase in the risk of developing childhood cancer [48–50]. Recent studies also supported an increased risk of pediatric cancer in individuals with birth defects unrelated to chromosomal abnormalities or known genetic syndromes [51, 52]. Hepatoblastoma, in particular, occurs in association with a wide variety of congenital abnormalities [53–55], especially craniosynostosis and renal anomalies [56], which we also observed in this study. In addition, a large case-control study confirmed the association between hepatoblastoma and kidney and bladder abnormalities [57]. In this study, we also described the cases of two patients with Hirschsprung disease and HB, a condition with extensive genetic heterogeneity [58, 59]; most Hirschsprung disease cases are sporadic (approximately 70%), while a smaller portion of patients have other associated congenital anomalies. However, pathogenic or likely pathogenic variants mapped to morbid OMIM genes that could explain the syndromic phenotypes of our patients were not detected, reinforcing that, despite the long-standing knowledge about the intersection between biological pathways of cancer and development, the etiology of most of these associations in specific cases remains unknown.

There is a recognized sexual dimorphism in hepatoblastoma, with increased prevalence in males [42, 60], and this study also reflected this tendency. We investigated whether this sex bias could be explained by an increased burden of rare damaging variants in one of the sexes; however, significant differences were not detected, suggesting that most likely environmental factors modulate the sexual dimorphism in HB prevalence.

In addition to SNV/indel variants, germline CNVs were already reported as causally related to cancer [61], but pathogenic CNVs affecting known/candidate CPGs were not detected in this study. One patient carried a gain of an entire Y chromosome, which is associated with an increased risk of learning disabilities, behavioral disturbances, and other clinical signs [62, 63]. Moreover, two 15q11.2 CNVs, considered risk factors for neurodevelopmental disorders [64, 65], were identified in two HB patients with global delay.

Germline mutations have been detected in 8-18% of children and adolescents with cancer [3, 8–10], and the most prevalent CPGs reported to be mutated are *TP53, APC, BRCA2, NF1, PMS2, RB1,* and*RUNX1* [3]. In a very recent work about the spectrum of germline mutations in childhood cancer, the majority of the mutations (55%) were mapped to genes not previously associated with the patient's tumor type [10]; this work included a small number of HB patients (3 patients) with no pathogenic or likely pathogenic variants. Our findings revealed a high burden of damaging germline variants mapped to CPGs, with only 40% of these patients presenting a family history of cancer. Most of the variants affected autosomal-dominant CPGs not previously linked to HB, since only one mutation was detected in a known gene of HB development

(*APC*). Interestingly, our data showed a clear predominance of CPGs related to gastrointestinal/colorectal cancer risk, such as familial adenomatous polyposis 1 and 2 (*APC* and*MUTYH*), nonpolyposis colorectal cancer hereditary types 1 (*MSH2*) and 6 (*TGFBR2*), Li-Fraumeni syndrome 2 (*CHEK2*), and Diamond-Blackfan anemia 1 (*RPS19*). Furthermore, the CPGs *DROSHA* and *VHL* , which showed P/LP variants, are known to be associated with renal cancer. *VHL* is a known predisposing gene for renal cell carcinoma, as well as other cancers, such as pheochromocytoma, hemangioblastoma, hypernephroma, pancreatic cancer, paraganglioma, and adenocarcinoma of the ampulla of Vater. Finally, a LoF variant was identified in the DNA repair gene*ERCC5* , in which homozygous mutations increase the skin cancer risk, and a damaging missense variant was mapped to *FAH* and causes an autosomal-recessive disorder characterized by progressive liver disease with increased liver cancer risk. Compared to previous pan-cancer studies, although not in HB patients, the findings of germline variants in *APC, CHEK2, ERCC5, MSH2, MUTYH,* and*VHL* were also detected in our work [3, 5, 10, 66].

We also observed numerous rare damaging VUS in known or candidate CPGs. Variants mapped to the *ATM, BRCA2, COL7A1, DHCR7, DOCK8, FANCD2, FANCM,* and *GLI3* genes were observed in more than one HB patient. In particular, the same *DHCR7* VUS (c.988G>A) was spotted in two patients (P09, born with only one functional kidney, and P26), both diagnosed with low-risk epithelial fetal hepatoblastoma; this gene encodes an enzyme that catalyzes the conversion of 7-dehydrocholesterol to cholesterol [67] and is a candidate for familial breast cancer in *BRCA1*- and *BRCA2* -negative breast cancer families [68], in a recessive mode of action. As in our study, VUS mapped to *MET, ATM, SLX4* , and *FANCA* were also reported in HB patients in a recently published large study [10].

Strong support exists for the hypothesis that monoallelic CPG mutations confer an increased risk of cancer for adult carriers, while biallelic carriers would have a high risk of childhood cancer [69]. More recently, large studies have shown that variants in heterozygosity affecting recessive genes can also increase the predisposition to pediatric cancer [70], and a second hit in the tumor, such as a loss of heterozygosity or an inactivation of the second allele, has been observed in some of these patients (4). A highlight has been given to the role of germline monoallelic variants in cancer predisposition in genes involved in the recognition and repair of DNA damage, such as the*ATM* , *PALB2* , and Fanconi anemia genes [71–77]. Since 1971 [78], it was proposed that individuals who were heterozygous for the Fanconi anemia genes might be at increased risk of cancer, and the premise would be that a modest reduction in the DNA repair efficiency could lead to tumor development. In our work, we detected P/LP monoallelic variants in CPGs associated with recessive clinical conditions (*FAH, ERCC5,* and *MUTYH*) and in DNA repair genes (*MSH2, CHEK2, ERCC5,* and *MUTYH*). In addition, we observed a general enrichment of heterozygous germline damaging VUS affecting DNA repair genes related to Fanconi anemia, nucleotide and base excision repair, and homologous recombination repair. Sequencing the tumors of the carriers could contribute to changing the classification of such variants [10] and clarifying the role of these variants in cancer predisposition.

One of the most interesting findings of this study was the disclosure of rare germline damaging variants affecting genes linked to liver differentiation and function, such as *FAH* , that cause a recessive disorder related to hepatocellular carcinoma risk. Several rare damaging variants were observed in genes of the cytochrome P450 (CYP) family, including one pathogenic heterozygous variant in *CYP21A2* (OMIM #201910 ADRENAL HYPERPLASIA, CONGENITAL, DUE TO 21-HYDROXYLASE DEFICIENCY, a recessive condition associated with testicular neoplasia in adults – P22) and two heterozygous pathogenic *CYP1B1* variants (OMIM #617315, #231300; recessive conditions) detected in different patients (P03 and P30). Ten rare variants mapped to *CYP1A1* (one LoF and nine missense variants), which is associated with primary liver metabolism, were present in eight patients (26%). *CYP1A1* encodes a xenobiotic-metabolizing enzyme acting in the placenta [79–81], as well as in several drugs and compounds widely used in pharmacotherapy [82, 83] or present in the diet [84]. *CYP1A1* expression is transcriptionally regulated through the AhR receptor [85–87] and various exogenous AhR binders, such as nitrosamines, polycyclic aromatic hydrocarbons, polychlorinated biphenyls, and halogenated dioxins, can be found in products from combustion processes, such as chimney soot, grilled food, and cigarette smoke, or as the product of incineration waste [88–91]. Elevated *CYP1A1* activity through AhR activation in the placentas of female smokers has been associated with

pregnancy complications, including low birth weight [92], a known risk factor for HB. In our previous study (14), in which we investigated the mutational burden of hepatoblastoma, a mutational signature similar to the COSMIC 29 signature was detected, which was initially observed only in oral gingivo-buccal squamous cell carcinoma, which develops in individuals with the habit of chewing tobacco. Currently, this mutational signature has also been detected in patients with lung, thyroid, bladder, biliary, breast, pancreas, liver, and kidney cancer and can be linked to nitrosamine exposure [93]. Therefore, we can speculate that a defective response during pregnancy to xenobiotic exposure, such as nitrosamines, could be linked to *CYP1A1*- damaging germline variants, increasing the risk for HB development and resulting in the mutational signature we previously found in HB.

In conclusion, our major findings in HB patients were the detection of germline variants in CPGs associated with gastrointestinal/colorectal and renal cancer risk, not always linked to a familial history of cancer; the enrichment of monoallelic variants in CPGs and DNA repair genes; and a high frequency of birth defects. Most studies of pediatric tumors are from North America, Europe, and Asia, and our study on HB is pioneering in South America and contributes to elucidating the genetic architecture of HB risk. Further validation of the genes highlighted as HB germline risk factors in other cohorts can provide new insights regarding HB development.

REFERENCES

1. Saletta F, Dalla Pozza L, Byrne JA. Genetic causes of cancer predisposition in children and adolescents. Transl Pediatr , 2015 4: 67–75.

2. Downing JR, Wilson RK, Zhang J, Mardis ER, Pui C-H, Ding L, Ley TJ, Evans WE. The Pediatric Cancer Genome Project. Nat Genet , 2012 44: 619–622.

3. Zhang J, Walsh MF, Wu G, Edmonson MN, Gruber TA, Easton J, Hedges D, Ma X, Zhou X, Yergeau DA, Wilkinson MR, Vadodaria B, Chen X, McGee RB, Hines-Dowell S, Nuccio R, Quinn E, Shurtleff SA, Rusch M, Patel A, Becksfort JB, Wang S, Weaver MS, Ding L, Mardis ER, Wilson RK, Gajjar A, Ellison DW, Pappo AS, Pui C-H, Nichols KE, Downing JR. Germline Mutations in Predisposition Genes in Pediatric Cancer. N Engl J Med , 2015 373: 2336–2346.

4. Kindler O, Quehenberger F, Benesch M, Seidel MG. The Iceberg Map of germline mutations in childhood cancer [Internet]. Curr Opin Pediatr , 2018 30: 855–863. [cited 2021 Jan 7] Available from: http://journals.lww.com/00008480-201812000-00023

5. Grobner SN, Worst BC, Weischenfeldt J, Buchhalter I, Kleinheinz K, Rudneva VA, Johann PD, Balasubramanian GP, Segura-Wang M, Brabetz S, Bender S, Hutter B, Sturm D, Pfaff E, Hubschmann D, Zipprich G, Heinold M, Eils J, Lawerenz C, Erkek S, Lambo S, Waszak S, Blattmann C, Borkhardt A, Kuhlen M, Eggert A, Fulda S, Gessler M, Wegert J, Kappler R, Baumhoer D, Burdach S, Kirschner-Schwabe R, Kontny U, Kulozik AE, Lohmann D, Hettmer S, Eckert C, Bielack S, Nathrath M, Niemeyer C, Richter GH, Schulte J, Siebert R, Westermann F, Molenaar JJ, Vassal G, Witt H, Burkhardt B, Kratz CP, Witt O, van Tilburg CM, Kramm CM, Fleischhack G, Dirksen U, Rutkowski S, Fruhwald M, von Hoff K, Wolf S, Klingebiel T, Koscielniak E, Landgraf P, Koster J, Resnick AC, Zhang J, Liu Y, Zhou X, Waanders AJ, Zwijnenburg DA, Raman P, Brors B, Weber UD, Northcott PA, Pajtler KW, Kool M, Piro RM, Korbel JO, Schlesner M, Eils R, Jones DTW, Lichter P, Chavez L, Zapatka M, Pfister SM. The landscape of genomic alterations across childhood cancers. Nature , 2018 555: 321–327.

6. Ma X, Liu Y, Liu Y, Alexandrov LB, Edmonson MN, Gawad C, Zhou X, Li Y, Rusch MC, John E, Huether R, Gonzalez-Pena V, Wilkinson MR, Hermida LC, Davis S, Sioson E, Pounds S, Cao X, Ries RE, Wang Z, Chen X, Dong L, Diskin SJ, Smith MA, Auvi JMG, Meltzer PS, Lau CC, Perlman EJ, Maris JM, Meshinchi S, Hunger SP, Gerhard DS, Zhang J. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. Nature , 2018.

7. Alejandro Sweet-Cordero E, Biegel JA. The genomic landscape of pediatric cancers: Implications for diagnosis and treatment. Science (80-) , 2019.

8. Akhavanfard S, Padmanabhan R, Yehia L, Cheng F, Eng C. Comprehensive germline genomic profiles of children, adolescents and young adults with solid tumors [Internet]. Nat Commun 2020 111 , 2020 11: 1–13. [cited 2021 Aug 17] Available from: https://www.nature.com/articles/s41467-020-16067-1

9. Capasso M, Montella A, Tirelli M, Maiorino T, Cantalupo S, Iolascon A. Genetic Predisposition to Solid Pediatric Cancers. Front Oncol , 2020 0: 2083.

10. Newman S, Nakitandwe J, Kesserwan CA, Azzato EM, Wheeler DA, Rusch M, Shurtleff S, Hedges DJ, Hamilton K V, Foy SG, Edmonson MN, Thrasher A, Bahrami A, Orr BA, Klco JM, Gu J, Harrison LW, Wang L, Clay MR, Ouma A, Silkov A, Liu Y, Zhang Z, Liu Y, Brady SW, Zhou X, Chang T-C, Pande M, Davis E, Becksfort J, Patel A, Wilkinson MR, Rahbarinia D, Kubal M, Maciaszek JL, Pastor V, Knight J, Gout AM, Wang J, Gu Z, Mullighan CG, Mcgee RB, Quinn EA, Nuccio R, Mostafavi R, Gerhardt EL, Taylor LM, Valdez JM, Hines-Dowell SJ, Pappo AS, Robinson G, Johnson L-M, Pui C-H, Ellison DW, Downing JR, Zhang J, Nichols KE. Genomes for Kids: The scope of pathogenic mutations in pediatric cancer revealed by comprehensive DNA and RNA sequencing, 2021.

11. Heck JE, Meyers TJ, Lombardi C, Park AS, Cockburn M, Reynolds P, Ritz B. Case-control study of birth characteristics and the risk of hepatoblastoma. Cancer Epidemiol , 2013 37: 390–395.

12. Stiller CA, Pritchard J, Steliarova-Foucher E. Liver cancer in European children: incidence and survival, 1978-1997. Report from the Automated Childhood Cancer Information System project. Eur J Cancer , 2006 42: 2115–2123.

13. Czauderna P, Garnier H. Hepatoblastoma: current understanding, recent advances, and controversies. F1000Research , 2018 7: 53.

14. Feng J, Polychronidis G, Heger U, Frongia G, Mehrabi A, Hoffmann K. Incidence trends and survival prediction of hepatoblastoma in children: A population-based study [Internet]. Cancer Commun , 2019 39: 62. [cited 2020 Jun 14] Available from: http://doi.wiley.com/10.1186/s40880-019-0411-7

15. Aguiar TFM, Rivas MP, Costa S, Maschietto M, Rodrigues T, Sobral de Barros J, Barbosa AC, Valieris R, Fernandes GR, Bertola DR, Cypriano M, Caminada de Toledo SR, Major A, Tojal I, Apezzato ML de P, Carraro DM, Rosenberg C, Lima da Costa CM, Cunha IW, Sarabia SF, Terrada D-L, Krepischi ACV. Insights Into the Somatic Mutation Burden of Hepatoblastomas From Brazilian Patients [Internet]. Front Oncol , 2020 10: 556. [cited 2020 May 5] Available from: https://www.frontiersin.org/article/10.3389/fonc.2020.00556/full

16. Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA). Incidência, mortalidade e morbidade hospitalar por câncer em crianças, adolescentes e adultos jovens no Brasil: informações dos registros de câncer e do sistema de mortalidade. Rio de Janeiro, INCA, 2016

17. Ries L, Smith M, Gurney J, Linet M, Tamra T, Young J, Bunin G. Cancer Incidence and Survival among Children and Adolescents: United States SEER Program 1975-1995, National Cancer Institute, SEER Program. Bethersda, MD, 1999

18. Turcotte LM, Georgieff MK, Ross JA, Feusner JH, Tomlinson GE, Malogolowkin MH, Krailo MD, Miller N, Fonstad R, Spector LG. Neonatal medical exposures and characteristics of low birth weight hepatoblastoma cases: a report from the Children's Oncology Group. Pediatr Blood Cancer , 2014 61: 2018–2023.

19. Pateva IB, Egler RA, Stearns DS. Hepatoblastoma in an 11-year-old: Case report and a review of the literature. Medicine (Baltimore) , 2017 96: e5858.

20. Tanimura M, Matsui I, Abe J, Ikeda H, Kobayashi N, Ohira M, Yokoyama M, Kaneko M. Increased risk of hepatoblastoma among immature children with a lower birth weight. Cancer Res , 1998 58: 3032–3035.

21. DeBaun MR, Tucker MA. Risk of cancer during the first four years of life in children from The Beckwith-Wiedemann Syndrome Registry. J Pediatr , 1998 132: 398–400.

22. Kim SY, Jung S-H, Kim MS, Han M-R, Park H-C, Jung ES, Lee SH, Lee SH, Chung Y-J. Genomic profiles of a hepatoblastoma from a patient with Beckwith-Wiedemann syndrome with uniparental disomy on chromosome 11p15 and germline mutation of APC and PALB2. Oncotarget , 2017.

23. Hirschman BA, Pollock BH, Tomlinson GE. The spectrum of APC mutations in children with hepatoblastoma from familial adenomatous polyposis kindreds. J Pediatr , 2005 147: 263–266.

24. Curia MC, Zuckermann M, De Lellis L, Catalano T, Lattanzio R, Aceto G, Veschi S, Cama A, Otte J-B, Piantelli M, Mariani-Costantini R, Cetta F, Battista P. Sporadic childhood hepatoblastomas show activation of beta-catenin, mismatch repair defects and p53 mutations. Mod Pathol , 2008 21: 7–14.

25. Kamien BA, Gabbett MT. Aicardi syndrome associated with hepatoblastoma and pulmonary sequestration. Am J Med Genet A , 2009 149A: 1850–1852.

26. Pereira EM, Marion R, Ramesh KH, Kim JS, Ewart M, Ricafort R. Hepatoblastoma in a mosaic trisomy 18 patient. J Pediatr Hematol Oncol , 2012 34: e145-8.

27. Sambrook J, Russell DW. Purification of nucleic acids by extraction with phenol:chloroform. CSH Protoc , 2006 2006.

28. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM [Internet]. 2013[cited 2020 May 18] Available from: http://github.com/lh3/bwa.

29. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res , 2010 20: 1297–1303.

30. Challis D, Yu J, Evani US, Jackson AR, Paithankar S, Coarfa C, Milosavljevic A, Gibbs RA, Yu F. An integrative variant analysis suite for whole exome next-generation sequencing data. BMC Bioinformatics , 2012 13.

31. Naslavsky MS, Scliar MO, Yamamoto GL, Wang JYT, Zverinova S, Karp T, Nunes K, Ceroni JRM, Carvalho DL de, Simões CE da S, Bozoklian D, Nonaka R, Silva N dos SB, Souza A da S, Andrade H de S, Passos MRS, Castro CFB, Mendes-Junior CT, Mercuri RL V., Miller TLA, Buzzo JL, Rego FO, Araújo NM, Magalhães WC, Mingroni-Netto RC, Borda V, Guio H, Barreto ML, Lima-Costa MF, Horta BL, Tarazona-Santos E, Meyer D, Galante PAF, Guryev V, Castelli EC, Duarte YAO, Passos-Bueno MR, Zatz M. Whole-genome sequencing of 1,171 elderly admixed individuals from the largest Latin American metropolis (São Paulo, Brazil) [Internet]. bioRxiv , 2020 10: 2020.09.15.298026. [cited 2021 Jul 7] Available from: https://www.biorxiv.org/content/10.1101/2020.09.15.298026v1

32. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, England EM, Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A, Seed C, Whiffin N, Chong JX, Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Minikel EV, Weisburd B, Lek M, Ware JS, Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly S, Ferriera S, Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R, Novod S, Petrillo N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K, Tolonen C, Wade G, Talkowski ME, Neale BM, Daly MJ, MacArthur DG. The mutational constraint spectrum quantified from variation in 141,456 humans [Internet]. Nat 2020 5817809 , 2020 581: 434–443. [cited 2021 Jul 7] Available from: https://www.nature.com/articles/s41586-020-2308-7

33. Clarke L, Fairley S, Zheng-Bradley X, Streeter I, Perry E, Lowy E, Tassé A-M, Flicek P. The international Genome sample resource (IGSR): A worldwide collection of genome variation incorporating the 1000 Genomes Project data [Internet]. Nucleic Acids Res , 2017 45: D854–D859. [cited 2021 Jul 7] Available from: https://academic.oup.com/nar/article/45/D1/D854/2770649

34. Fuentes Fajardo K V, Adams D, Mason CE, Sincan M, Tifft C, Toro C, Boerkoel CF, Gahl W, Markello T. Detecting false-positive signals in exome sequencing. Hum Mutat , 2012 33: 609–613.

35. Stelzer G, Plaschkes I, Oz-Levi D, Alkelai A, Olender T, Zimmerman S, Twik M, Belinky F, Fishilevich S, Nudel R, Guan-Golan Y, Warshawsky D, Dahary D, Kohn A, Mazor Y, Kaplan S, Iny Stein T, Baris HN, Rappaport N, Safran M, Lancet D. VarElect: The phenotype-based variation prioritizer of the GeneCards Suite. BMC Genomics , 2016.

36. Köhler S, Carmody L, Vasilevsky N, Jacobsen JOB, Danis D, Gourdine JP, Gargano M, Harris NL, Matentzoglu N, McMurry JA, Osumi-Sutherland D, Cipriani V, Balhoff JP, Conlin T, Blau H, Baynam G, Palmer R, Gratian D, Dawkins H, Segal M, Jansen AC, Muaz A, Chang WH, Bergerson J, Laulederkind SJF, Yüksel Z, Beltran S, Freeman AF, Sergouniotis PI, Durkin D, Storm AL, Hanauer M, Brudno M, Bello SM, Sincan M, Rageth K, Wheeler MT, Oegema R, Lourghi H, Della Rocca MG, Thompson R, Castellanos F, Priest J, Cunningham-Rundles C, Hegde A, Lovering RC, Hajek C, Olry A, Notarangelo L, Similuk M, Zhang XA, Gómez-Andrés D, Lochmüller H, Dollfus H, Rosenzweig S, Marwaha S, Rath A, Sullivan K, Smith C, Milner JD, Leroux D, Boerkoel CF, Klion A, Carter MC, Groza T, Smedley D, Haendel MA, Mungall C, Robinson PN. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. Nucleic Acids Res , 2019.

37. Kopanos C, Tsiolkas V, Kouris A, Chapple CE, Albarca Aguilera M, Meyer R, Massouras A. VarSome: the human genomic variant search engine. Bioinformatics , 2019.

38. Hampel H, Bennett RL, Buchanan A, Pearlman R, Wiesner GL, Guideline Development Group, American College of Medical Genetics and Genomics Professional Practice and Guidelines Committee and National Society of Genetic Counselors Practice Guidelines Committee. A practice guideline from the American College of Medical Genetics and Genomics and the National Society of Genetic Counselors: referral indications for cancer predisposition assessment. [Internet]. Genet Med , 2015 17: 70–87. [cited 2019 Aug 4] Available from: http://www.ncbi.nlm.nih.gov/pubmed/25394175

39. LSue Richards, PhD1 , Nazneen Aziz, PhD2, 16 , Sherri Bale, PhD3 , David Bick M, , Soma Das P, Julie Gastier-Foster, PhD6, 7, 8, Wayne W. Grody, MD, PhD9, 10, 11, Madhuri Hegde P, Elaine Lyon P, , Elaine Spector P, , Karl Voelkerding M, and Heidi L. Rehm P, On ;, Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL, Laboratories KD, Genetics M, Health O, Road P, Molecular C, Children N, State O, Berindan-neagoe I, Monroig P, Pasculli B, George A, Medicine T, Hatieganu PI, Juan S, Rico P, Sciences P, Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL. Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology Sue. Genet Med , 2015.

40. Oscanoa J, Sivapalan L, Gadaleta E, Dayem Ullah AZ, Lemoine NR, Chelala C. SNPnexus: a web server for functional annotation of human genome sequence variation (2020 update). Nucleic Acids Res , 2020 48: W185–W192.

41. Oliveira D, Leal GF, Sertie AL, Caires LCJ, Goulart E, Musso CM, Oliveira JRM de, Krepischi ACV, Vianna-Morgante AM, Zatz M. 10q23.31 microduplication encompassing PTEN decreases mTOR signalling activity and is associated with autosomal dominant primary microcephaly. J Med Genet , 2018.

42. Spector LG, Birch J. The epidemiology of hepatoblastoma. Pediatr Blood Cancer , 2012.

43. Czauderna P, Haeberle B, Hiyama E, Rangaswami A, Krailo M, Maibach R, Rinaldi E, Feng Y, Aronson D, Malogolowkin M, Yoshimura K, Leuschner I, Lopez-Terrada D, Hishiki T, Perilongo G, von Schweinitz D, Schmid I, Watanabe K, Derosa M, Meyers R. The Children's Hepatic tumors International Collaboration (CHIC): Novel global rare tumor database yields new prognostic factors in hepatoblastoma and becomes a research model. Eur J Cancer , 2016 52: 92–101.

44. Meyers RL, Maibach R, Hiyama E, Haberle B, Krailo M, Rangaswami A, Aronson DC, Malogolowkin MH, Perilongo G, von Schweinitz D, Ansari M, Lopez-Terrada D, Tanaka Y, Alaggio R, Leuschner I, Hishiki T, Schmid I, Watanabe K, Yoshimura K, Feng Y, Rinaldi E, Saraceno D, Derosa M, Czauderna P. Risk-stratified staging in paediatric hepatoblastoma: a unified analysis from the Children's Hepatic tumors International Collaboration. Lancet Oncol , 2017 18: 122–131.

45. Pinto RB, Ramos ARL, Backes AN, Santos BJ Dos, Provenzi VO, Carbonera MR, Roenick ML, Santos PPA Dos, Falhauber F, Souza MV de, Bassols JV, Artigalas O. Hirschsprung disease and hepatoblastoma: case report of a rare association. Sao Paulo Med J , 2016 134: 171–175.

46. Rahman N. Realizing the promise of cancer predisposition genes. Nature , 2014 505: 302–308.

47. Foulkes WD, Priest JR, Duchaine TF. DICER1: mutations, microRNAs and mechanisms. Nat Rev Cancer , 2014 14: 662–672.

48. Johnson KJ, Lee JM, Ahsan K, Padda H, Feng Q, Partap S, Fowler SA, Druley TE. Pediatric cancer risk in association with birth defects: A systematic review. PLoS One , 2017 12: e0181246.

49. Krivit W, Good RA. Simultaneous Occurrence of Mongolism and Leukemia: Report of a Nationwide Survey. AMA J Dis Child , 1957.

50. Agha MM, Williams JI, Marrett L, To T, Zipursky A, Dodds L. Congenital abnormalities and childhood cancer. Cancer , 2005 103: 1939–1948.

51. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2017. CA Cancer J Clin , 2017 67: 7–30.

52. Norwood MS, Lupo PJ, Chow EJ, Scheurer ME, Plon SE, Danysh HE, Spector LG, Carozza SE, Doody DR, Mueller BA. Childhood cancer risk in those with chromosomal and non-chromosomal congenital anomalies in Washington State: 1984-2013. PLoS One , 2017 12: e0179006.

53. Scollon S, Anglin AK, Thomas M, Turner JT, Wolfe Schneider K. A Comprehensive Review of Pediatric Tumors and Associated Cancer Predisposition Syndromes. J Genet Couns , 2017 26: 387–434.

54. Ansell P, Mitchell CD, Roman E, Simpson J, Birch JM, Eden TOB. Relationships between perinatal and maternal characteristics and hepatoblastoma: a report from the UKCCS. Eur J Cancer , 2005 41: 741–748.

55. Narod SA, Hawkins MM, Robertson CM, Stiller CA. Congenital anomalies and childhood cancer in Great Britain. Am J Hum Genet , 1997 60: 474–485.

56. de Camargo B, de Oliveira Ferreira JM, de Souza Reis R, Ferman S, de Oliveira Santos M, Pombo-de-Oliveira MS. Socioeconomic status and the incidence of non-central nervous system childhood embryonic tumours in Brazil. BMC Cancer , 2011.

57. Venkatramani R, Spector LG, Georgieff M, Tomlinson G, Krailo M, Malogolowkin M, Kohlmann W, Curtin K, Fonstad RK, Schiffman JD. Congenital abnormalities and hepatoblastoma: a report from the Children's Oncology Group (COG) and the Utah Population Database (UPDB). Am J Med Genet A , 2014 164A: 2250–2255.

58. Sergi CM, Caluseriu O, McColl H, Eisenstat DD. Hirschsprung's disease: clinical dysmorphology, genes, micro-RNAs, and future perspectives. Pediatr Res , 2017 81: 177–191.

59. Heuckeroth RO. Hirschsprung disease - integrating basic science and clinical medicine to improve outcomes. Nat Rev Gastroenterol Hepatol , 2018 15: 152–167.

60. Williams LA, Spector LG. Survival Differences Between Males and Females Diagnosed With Childhood Cancer. JNCI cancer Spectr , 2019 3: pkz032.

61. Krepischi ACV, Pearson PL, Rosenberg C. Germline copy number variations and cancer predisposition. Future Oncol , 2012 8: 441–450.

62. Lalatta F, Folliero E, Cavallari U, Segni M Di, Gentilin B, Fogliani R, Quagliarini D, Vizziello P, Monti F, Gargantini L. Early manifestations in a cohort of children prenatally diagnosed with 47,XYY. Role of multidisciplinary counseling for parental guidance and prevention of aggressive behavior [Internet]. Ital J Pediatr , 2012 38: 52. [cited 2021 Aug 18] Available from: /pmc/articles/PMC3523010/

63. Bardsley MZ, Kowal K, Levy C, Gosek A, Ayari N, Tartaglia N, Lahlou N, Winder B, Grimes S, Ross JL. 47,XYY Syndrome: Clinical Phenotype and Timing of Ascertainment [Internet]. J Pediatr , 2013 163: 1085. [cited 2021 Sep 28] Available from: /pmc/articles/PMC4097881/

64. Butler MG. Clinical and genetic aspects of the 15q11.2 BP1-BP2 microdeletion disorder. J Intellect Disabil Res , 2017 61: 568–579.

65. Rafi SK, Butler MG. The 15q11.2 BP1-BP2 Microdeletion (Burnside-Butler) Syndrome: In Silico Analyses of the Four Coding Genes Reveal Functional Associations with Neurodevelopmental Phenotypes. Int J Mol Sci , 2020 21.

66. Ma X, Liu Y, Liu Y, Alexandrov LB, Edmonson MN, Gawad C, Zhou X, Li Y, Rusch MC, John E, Huether R, Gonzalez-Pena V, Wilkinson MR, Hermida LC, Davis S, Sioson E, Pounds S, Cao X, Ries RE, Wang Z, Chen X, Dong L, Diskin SJ, Smith MA, Auvi JMG, Meltzer PS, Lau CC, Perlman EJ, Maris JM, Meshinchi S, Hunger SP, Gerhard DS, Zhang J. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. Nature , 2018 555: 371–376.

67. Fitzky BU, Witsch-Baumgartner M, Erdel M, Lee JN, Paik Y-K, Glossmann H, Utermann G, Moebius FF. Mutations in the Δ7-sterol reductase gene in patients with the Smith–Lemli–Opitz syndrome [Internet]. Proc Natl Acad Sci , 1998 95: 8181–8186. [cited 2021 Jul 7] Available from: https://www.pnas.org/content/95/14/8181

68. Shahi RB, De Brakeleer S, Caljon B, Pauwels I, Bonduelle M, Joris S, Fontaine C, Vanhoeij M, Van Dooren S, Teugels E, De Grève J. Identification of candidate cancer predisposing variants by performing whole-exome sequencing on index patients from BRCA1 and BRCA2-negative breast cancer families. BMC Cancer , 2019 19: 313.

69. Rahman N, Scott RH. Cancer genes associated with phenotypes in monoallelic and biallelic mutation carriers: New lessons from old players. Hum Mol Genet , 2007.

70. Savary C, Kim A, Lespagnol A, Gandemer V, Pellier I, Andrieu C, Pagès G, Galibert MD, Blum Y, de Tayrac M. Depicting the genetic architecture of pediatric cancers through an integrative gene network approach. Sci Rep , 2020 10: 1–15.

71. Thompson D, Duedal S, Kirner J, McGuffog L, Last J, Reiman A, Byrd P, Taylor M, Easton DF. Cancer risks and mortality in heterozygous ATM mutation carriers. J Natl Cancer Inst , 2005 97: 813–822.

72. Renwick A, Thompson D, Seal S, Kelly P, Chagtai T, Ahmed M, North B, Jayatilake H, Barfoot R, Spanova K, McGuffog L, Evans DG, Eccles D, Easton DF, Stratton MR, Rahman N. ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. Nat Genet , 2006 38: 873–875.

73. Antoniou AC, Casadei S, Heikkinen T, Barrowdale D, Pylkas K, Roberts J, Lee A, Subramanian D, De Leeneer K, Fostira F, Tomiak E, Neuhausen SL, Teo ZL, Khan S, Aittomaki K, Moilanen JS, Turnbull C, Seal S, Mannermaa A, Kallioniemi A, Lindeman GJ, Buys SS, Andrulis IL, Radice P, Tondini C, Manoukian S, Toland AE, Miron P, Weitzel JN, Domchek SM, Poppe B, Claes KBM, Yannoukakos D, Concannon P, Bernstein JL, James PA, Easton DF, Goldgar DE, Hopper JL, Rahman N, Peterlongo P, Nevanlinna H, King M-C, Couch FJ, Southey MC, Winqvist R, Foulkes WD, Tischkowitz M. Breast-cancer risk in families with mutations in PALB2. N Engl J Med , 2014 371: 497–506.

74. Helgason H, Rafnar T, Olafsdottir HS, Jonasson JG, Sigurdsson A, Stacey SN, Jonasdottir A, Tryggvadottir L, Alexiusdottir K, Haraldsson A, le Roux L, Gudmundsson J, Johannsdottir H, Oddsson A, Gylfason A, Magnusson OT, Masson G, Jonsson T, Skuladottir H, Gudbjartsson DF, Thorsteinsdottir U, Sulem P,

Stefansson K. Loss-of-function variants in ATM confer risk of gastric cancer [Internet]. Nat Genet , 2015 47: 906–910. [cited 2020 Feb 29] Available from: http://www.nature.com/articles/ng.3342

75. Esteban-Jurado C, Franch-Expósito S, Muñoz J, Ocaña T, Carballal S, López-Cerón M, Cuatrecasas M, Vila-Casadesús M, Lozano JJ, Serra E, Beltran S, Brea-Fernández A, Ruiz-Ponte C, Castells A, Bujanda L, Garre P, Caldés T, Cubiella J, Balaguer F, Castellví-Bel S. The Fanconi anemia DNA damage repair pathway in the spotlight for germline predisposition to colorectal cancer [Internet]. Eur J Hum Genet , 2016 24: 1501–1505. [cited 2020 Feb 29] Available from: http://www.nature.com/articles/ejhg201644

76. Esai Selvan M, Klein RJ, Gümüş ZH. Rare, Pathogenic Germline Variants in *Fanconi Anemia* Genes Increase Risk for Squamous Lung Cancer [Internet]. Clin Cancer Res , 2019 25: 1517–1525. [cited 2020 Feb 29] Available from: http://clincancerres.aacrjournals.org/lookup/doi/10.1158/1078-0432.CCR-18-2660

77. Rebbeck TR, Mitra N, Wan F, Sinilnikova OM, Healey S, McGuffog L, Mazoyer S, Chenevix-Trench G, Easton DF, Antoniou AC, Nathanson KL, Laitman Y, Kushnir A, Paluch-Shimon S, Berger R, Zidan J, Friedman E, Ehrencrona H, Stenmark-Askmalm M, Einbeigi Z, Loman N, Harbst K, Rantala J, Melin B, Huo D, Olopade OI, Seldon J, Ganz PA, Nussbaum RL, Chan SB, Odunsi K, Gayther SA, Domchek SM, Arun BK, Lu KH, Mitchell G, Karlan BY, Walsh C, Lester J, Godwin AK, Pathak H, Ross E, Daly MB, Whittemore AS, John EM, Miron A, Terry MB, Chung WK, Goldgar DE, Buys SS, Janavicius R, Tihomirova L, Tung N, Dorfling CM, van Rensburg EJ, Steele L, Neuhausen SL, Ding YC, Ejlertsen B, Gerdes A-M, Hansen T v O, Ramon y Cajal T, Osorio A, Benitez J, Godino J, Tejada M-I, Duran M, Weitzel JN, Bobolis KA, Sand SR, Fontaine A, Savarese A, Pasini B, Peissel B, Bonanni B, Zaffaroni D, Vignolo-Lutati F, Scuvera G, Giannini G, Bernard L, Genuardi M, Radice P, Dolcetti R, Manoukian S, Pensotti V, Gismondi V, Yannoukakos D, Fostira F, Garber J, Torres D, Rashid MU, Hamann U, Peock S, Frost D, Platte R, Evans DG, Eeles R, Davidson R, Eccles D, et al. Association of type and location of BRCA1 and BRCA2 mutations with risk of breast and ovarian cancer. JAMA , 2015 313: 1347–1361.

78. Swift M, Zimmerman D, McDonough ER. Squamous cell carcinomas in Fanconi's anemia. JAMA , 1971 216: 325–326.

79. Suter M, Abramovici A, Showalter L, Hu M, Shope C Do, Varner M, Aagaard-Tillery K. IN UTERO TOBACCO EXPOSURE EPIGENETICALLY MODIFIES PLACENTAL CYP1A1 EXPRESSION [Internet]. Metabolism , 2010 59: 1481. [cited 2021 Jul 7] Available from: /pmc/articles/PMC2921565/

80. Yan Y, Wang H, Feng Y. Alterations of placental cytochrome P450 1A1 and P-glycoprotein in tobacco-induced intrauterine growth retardation in rats [Internet]. Acta Pharmacol Sin 2005 2611 , 2005 26: 1387–1394. [cited 2021 Jul 7] Available from: https://www.nature.com/articles/aps2005203

81. Stejskalova L, Pavek P. The function of cytochrome P450 1A1 enzyme (CYP1A1) and aryl hydrocarbon receptor (AhR) in the placenta. Curr Pharm Biotechnol , 2011 12: 715–730.

82. Kapelyukh Y, Henderson CJ, Scheer N, Rode A, Wolf CR. Defining the Contribution of CYP1A1 and CYP1A2 to Drug Metabolism Using Humanized CYP1A1/1A2 and Cyp1a1/Cyp1a2 Knockout Mice [Internet]. Drug Metab Dispos , 2019 47: 907. [cited 2021 Jul 7] Available from: /pmc/articles/PMC6657216/

83. Willis AJ, Indra R, Wohak LE, Sozeri O, Feser K, Mrizova I, Phillips DH, Stiborova M, Arlt VM. The impact of chemotherapeutic drugs on the CYP1A1-catalysed metabolism of the environmental carcinogen benzo[a]pyrene: Effects in human colorectal HCT116 TP53(+/+), TP53(+/-) and TP53(-/-) cells. Toxicology , 2018 398–399: 1–12.

84. Ito S, Chen C, Satoh J, Yim S, Gonzalez FJ. Dietary phytochemicals regulate whole-body CYP1A1 expression through an arylhydrocarbon receptor nuclear translocator–dependent system in gut [Internet]. J Clin Invest , 2007 117: 1940. [cited 2021 Jul 7] Available from: /pmc/articles/PMC1890999/

85. Delescluse C, Lemaire G, de Sousa G, Rahmani R. Is CYP1A1 induction always related to AHR signaling pathway? [Internet]. Toxicology , 2000 153: 73–82. [cited 2020 Mar 31] Available from: http://www.ncbi.nlm.nih.gov/pubmed/11090948

86. Kawashiro T, Yamashita K, Zhao XJ, Koyama E, Tani M, Chiba AN, Ishizaki T. A study on the metabolism of etoposide and possible interactions with antitumor or supporting Agents by Human Liver Microsomes. J Pharmacol Exp Ther , 1998 286: 1294–1300.

87. Kamenickova A, Anzenbacherova E, Pavek P, Soshilov AA, Denison MS, Zapletalova M, Anzenbacher P, Dvorak Z. Effects of anthocyanins on the AhR-CYP1A1 signaling pathway in human hepatocytes and human cancer cell lines [Internet]. Toxicol Lett , 2013 221: 1. [cited 2021 Sep 28] Available from: /pmc/articles/PMC3759295/

88. Shahin NN, Abd-Elwahab GT, Tawfiq AA, Abdelgawad HM. Potential role of aryl hydrocarbon receptor signaling in childhood obesity. Biochim Biophys Acta - Mol Cell Biol Lipids , 2020 1865: 158714.

89. Schanz O, Chijiiwa R, Cengiz SC, Majlesain Y, Weighardt H, Takeyama H, Förster I. Dietary AhR Ligands Regulate AhRR Expression in Intestinal Immune Cells and Intestinal Microbiota Composition [Internet]. Int J Mol Sci , 2020 21. [cited 2021 Jul 7] Available from: /pmc/articles/PMC7246938/

90. Lu J, Shang X, Zhong W, Xu Y, Shi R, Wang X. New insights of CYP1A in endogenous metabolism: a focus on single nucleotide polymorphisms and diseases. Acta Pharm Sin B , 2020 10: 91–104.

91. Larigot L, Juricek L, Dairou J, Coumoul X. AhR signaling pathways and regulatory functions. Biochim Open , 2018 7: 1–9.

92. Stejskalova L, Dvorak Z, Pavek P. Endogenous and Exogenous Ligands of Aryl Hydrocarbon Receptor: Current State of Art. 2011

93. Alexandrov LB, Jones PH, Wedge DC, Sale JE, Campbell PJ, Nik-Zainal S, Stratton MR. Clock-like mutational processes in human somatic cells. Nat Genet , 2015 47: 1402–1407.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author/s.

COMPLIANCE WITH ETHICAL COMMITTEE

Thirty children diagnosed with hepatoblastoma were enrolled in this study, and their parents, when available. Patients are from five different Brazilian‘s institutes: A. C. Camargo Cancer Center, Adolescent and Child with Cancer Support Group (GRAACC), Pediatric Cancer Institute (ITACI), Baleia‘s Hospital and Hospital da Criança Conceição. The Research Ethics Committee of the respective Institutions approved this research using these biological samples, and all samples were collected after informed signed consent was obtained from parents or legal guardians.

AUTHOR CONTRIBUTIONS

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Talita Aguiar, Anne Teixeira, Juliana Sobral and Ana Krepischi. The first draft of the manuscript was written by Talita Aguiar, Anne Teixeira, and Ana Krepischi and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

CONFLICTS OF INTEREST

We declare that we have no conflicts of interest.

ACKNOWLEDGMENTS

We thank the patients and their families who enrolled in this study.

FUNDING

The present study was supported by grants from FAPESP (CEPID - Human Genome and Stem Cell Research Center 2013/08028-1; 2018/21047-9; fellowships: 2018/05961-2, 2016/04785-0, 2019/17423-8) and CNPq (141625/2016-3). The funders had no roles in the study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

LEGENDS OF FIGURES

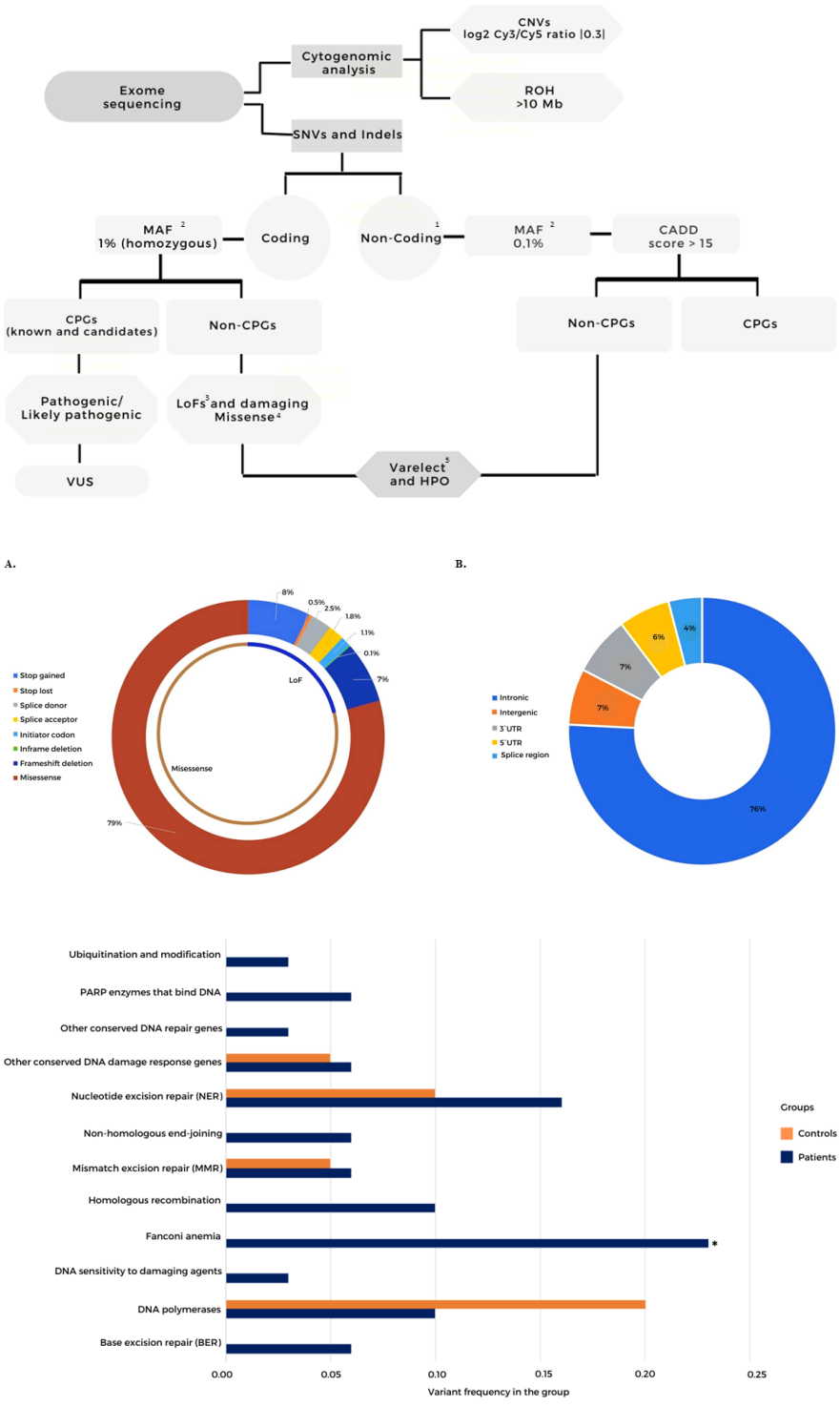
Figure 1: Workflow of the exome sequencing analysis. Variants were filtered according to quality (Phred score >20, read depth >10, variant allele frequency >35%); coding and noncoding variants were separately analyzed; frequency (GnomAD, ABraOM, 1K Genomes); the effect on coding variants (frameshift, stop loss/gain, missense, splice site, nonsense); for missense variants, the prediction of pathogenicity in at least 5 out of 6 algorithms; HPO annotation (hepatoblastoma, abnormalities of the liver, and cancer). The filtered variants were visually examined using Integrative Genomics Viewer (IGV) software (http//www.broadinstitute.org/igv) to further filter out possible strand bias and homopolymeric region artifacts. All the filtered variants mapped to cancer predisposition genes were annotated using the ACMG guidelines. 1- Intronic variants, 3'UTR, 5'UTR; 2- MAF: GnomAD, ABraOM, 1K Genomes; 3- Frameshift, stop loss, stop gain, missense, splice site, nonsense variants; 4- Missense variants with dbNSFP Functional Prediction of pathogenicity in at least 5 out of 6 algorithms; 5- Terms used for Varelect and HPO annotation: Hepatoblastoma, abnormalities of the liver, and cancer. CNV - copy number variation, ROH - region of homozygosity, MAF – maximum allele frequency, CPG – cancer predisposition gene, VUS – variant of uncertain significance, HPO - Human Phenotype Ontology, HB - hepatoblastoma.

Figure 2: Distribution of the detected high-quality rare coding and noncoding variants detected in 30 HB patients.

Sequence ontology of the rare coding variants detected after selection by the read depth (>10), Phred score (>20), alternative allele frequency (>0.35), and population frequency (<1%). A total of 2107 variants were classified into 1671 missense mutations and 436 LoF variants.

Sequence ontology of the rare noncoding variants detected after selection by the read depth (>10), Phred score (>20), alternative allele frequency (>0.35), and population frequency (<0.1%). A total of 2,070 noncoding variants were distributed in intronic, intergenic, and 3' and 5' UTRs.

Figure 3 : **Frequency of high-quality rare germline coding variants mapped to DNA repair genes in HB patients and a control group.** A list of 220 DNA repair genes distributed in 16 categories was analyzed; the 12 categories with variants detected in either patients or controls are represented. PARP – poly (ADP-ribose) polymerase. * p value 0.0338; Fisher's test.



Hosted file

Tables.docx available at https://authorea.com/users/328345/articles/545133-unraveling-the-genetic-architecture-of-hepatoblastoma-risk-birth-defects-and-increased-burden-

of-germline-damaging-variants-in-gastrointestinal-renal-cancer-predisposition-and-dna-repair-genes