

HABP2 p.G534E variant in patients with family history of thyroid and breast cancer

Maisa Pinheiro^{1,2}, Sandra Aparecida Drigo², Renata Tonhosolo¹, Sonia C.S. Andrade³, Fabio Albuquerque Marchi¹, Igor Jurisica^{4,5}, Luiz Paulo Kowalski⁶, Maria Isabel Achatz^{1,7} and Silvia Regina Rogatto^{1,2,8}

¹CIPE - International Research Center, A. C. Camargo Cancer Center, Sao Paulo, SP, Brazil

²Department of Urology, Faculty of Medicine, São Paulo State University, UNESP, Botucatu, SP, Brazil

³Department of Genetics and Evolutionary Biology, University of Sao Paulo, USP, Sao Paulo, SP, Brazil

⁴Princess Margaret Cancer Centre, University Health Network and The University of Toronto, Toronto, ON, Canada

⁵Institute of Neuroimmunology, Slovak Academy of Sciences, Bratislava, Slovakia

⁶Department of Head and Neck Surgery and Otorhinolaryngology, A. C. Camargo Cancer Center, Sao Paulo, SP, Brazil

⁷Division of Cancer Epidemiology and Genetics, National Cancer Institute/National Institutes of Health, Bethesda, MD, USA

⁸Department of Clinical Genetics, Vejle Hospital, Institute of Regional Health Research, University of Southern Denmark, Vejle, Denmark

Correspondence to: Silvia Regina Rogatto, **email:** silvia.regina.rogatto@rsyd.dk, rogatto@fmb.unesp.br

Keywords: breast cancer, thyroid cancer, genetics, molecular markers, hereditary tumors

Received: February 07, 2017

Accepted: March 13, 2017

Published: March 29, 2017

Copyright: Pinheiro et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Familial Papillary Thyroid Carcinoma (PTC) has been described as a hereditary predisposition cancer syndrome associated with mutations in candidate genes including HABP2. Two of 20 probands from families with history of PTC and breast carcinoma (BC) were evaluated by whole exome sequencing (WES) revealing HABP2 p.G534E. Sanger sequencing was used to confirm the involvement of this variant in three families (F1: 7 relatives; F2: 3 and F3: 3). The proband and his sister (with no malignant tumor so far) from F1 were homozygous for the variant whereas one relative with PTC from F2 was negative for the variant. Although the proband of the F3 with PTC was HABP2 wild type, three relatives presented the variant. Five of 170 healthy Brazilian individuals with no family history of BC or PTC and three of 50 sporadic PTC presented the p.G534E. These findings suggested no association of this variant with our familial PTC cases. Genes potentially associated with deregulation of the extracellular matrix organization pathway (CTSB, TNXB, COL4A3, COL16A1, COL24A1, COL5A2, NID1, LOXL2, MMP11, TRIM24 and MUSK) and DNA repair function (NBN and MSH2) were detected by WES, suggesting that other cancer-associated genes have pathogenic effects in the risk of familial PTC development.

INTRODUCTION

Thyroid Carcinoma (TC) is a common endocrine neoplasia corresponding to 1% of all cancers worldwide [1, 2]. Non-Medullary Thyroid Carcinoma (NMTC), which includes Papillary Thyroid Carcinoma (PTC) and Follicular Thyroid Carcinoma (FTC), is the most common tumor involving the thyroid gland, the former being responsible for 80-95% of all cases [3, 4].

Family history of NMTCs has been associated with well known Hereditary Cancer Predisposition Syndromes (HCPS) including Cowden, Familial Adenomatous Polyposis/Gardner, Carney Complex type 1, Werner and DICER1 syndromes, which were associated with mutations in *PTEN*, *APC*, *PRKARIA*, *WRN* and *DICER1* genes, respectively [5, 6]. Although TC has been reported in families with Lynch and Li-Fraumeni syndromes, its inclusion in the spectrum of tumors of these disorders remains to be clarified [7-11].

The predominance of PTC in families having first or second degree relatives affected by the disease has been described as a particular clinical entity suggestive of a heritable component. Based on this premise, new candidate genes have been reported as associated with familial PTC including *HABP2*, *SRGAP1*, *PARP4* and *SRRM2* [12-15].

Gara et al. (2015) reported a novel non-synonymous p.G534E variant in the Hyaluronan Binding Protein 2 gene (*HABP2*, which encodes for a member of the peptidase S1 family of serine proteases) associated with familial NMTC. Based on the loss of function, the authors suggested that *HABP2* is a tumor suppressor gene with dominant-negative effect [15]. Zhang and Xing reported the presence of this variant in heterozygous form in six individuals with PTC from four kindred [16]. Contrary to these reports, Tomsic et al. (2016) evaluated the variant *HABP2* p.G534E in 179 individuals with family history of NMTC, 1160 sporadic PTC and 1395 controls showing an allele frequency of 0.61, 0.8 and 0.87, respectively. In six of eight families, the authors did not find co-segregation of the mutated allele and phenotype [17].

An intensive debate has been raised in literature [18-20] based on the allele frequency described as common in several populations as well as in public and in house databases. In highly inbred Middle Eastern population, Alzahrani et al. (2016) described the absence of mutations in familial NMTC and only one of 509 sporadic cases harbored the variant [21]. In Hispanic individuals (281 NMTC and 1105 population-matched controls) Bohórquez et al. (2016) reported that the *HABP2* p.G534E was not significantly associated with cancer susceptibility [22]. In British population, no significant association was found between the p.G534E and NMTC [23]. Similarly, the variant was not associated with PTC neither in Spanish nor Australian families [24, 25]. According to Carvajal-Carmona et al. (2016), the high allele frequency is a plausible explanation for the identification of carriers in different populations [26].

In this study we identified the *HABP2* p.G534E mutation in three Brazilian families with history of PTC and BC. We also assessed the presence of the variant in sporadic PTC cases as well as in healthy Brazilian individuals. Extracellular matrix organization pathway and DNA repair function might be potentially deregulated and contribute to the risk of developing thyroid familial cancer.

RESULTS

Detection of *HABP2* p.G534E variant in patients with familial history of NMTC and BC

Family 1

The male patient diagnosed with PTC (Case 1: IV-2) was detected as homozygous for the *HABP2* p.G534E

variant by exome sequencing analysis and confirmed by Sanger sequencing. He had no clinical or pathological features of tumor aggressiveness (Figure 1, Supplementary Table 1). Seven family members (4 affected and 3 unaffected) were screened by Sanger sequencing revealing six positive individuals for the *HABP2* p.G534E (Figure 1, Supplementary Table 1). The sister of the patient (IV-1 with history of thyroid goiter and benign endometrial polyp) was homozygous for the mutation (Figure 1). The variant in heterozygous form was detected in his mother with skin carcinoma (III-5), the sister diagnosed with PTC (IV-5), two relatives with PTC (IV-5 and V-1) and two relatives unaffected by cancer (IV-3 and V-2) (Figure 1). His cousin diagnosed with BC (IV-6) was negative for p.G534E (Figure 1B).

Family 2

A woman with history of invasive ductal BC (Case 2; III-12) presented the *HABP2* p.G534E variant detected by WES. She was also diagnosed with four different benign tumors in a period of six years, including thyroid goiter, breast nodules, skin nodules and colorectal polyps. Sanger sequencing confirmed the presence of this variant in the index case and in two of three sisters (III-5: thyroid colloid goiter and colorectal hyperplastic polyp; III-6: thyroid follicular adenoma and benign breast nodule) (Figure 1A). The only sister diagnosed with PTC (III-9) was negative for p.G534E (Figure 1B, Supplementary Table 1).

Family 3

Sanger sequencing of *HABP2* exon 13 was carried out in the index patient (Case 3) and three relatives from family 3. The 40 year-old female proband (III-5) diagnosed with PTC was negative for p.G534E mutation. Three relatives tested were positive for the same variant (Figure 1, Supplementary Table 1) including the mother (II-4: PTC), the female cousin (III-8: PTC) and the proband's niece (IV-3: no history of cancer). The index patient had also breast cancer family history from maternal lineage (II-7) (Figure 1). Despite not being possible to assess her aunt (II-5), diagnosed with PTC and colorectal carcinoma, we may conclude she is positive for p.G534E since her daughter (III-9) carried this allele (Figure 1).

Sporadic PTC and healthy Brazilian individuals

The DNA of blood samples from both, 170 healthy individuals and 50 patients with sporadic PTC, were screened by Sanger sequencing to investigate the *HABP2* exon 13. Five healthy individuals were heterozygous for the variant, which represents an Allele Frequency (AF) of 0.0147. In PTC samples the AF was 0.04: two tumors were heterozygous and one was homozygous for p.G534E. Clinical features from sporadic PTC are presented in Supplementary Table 2.

Variant prioritization and protein-protein interaction network

Our results suggested no co-segregation of the *HABP2* p.G534E mutated allele and the phenotype. Furthermore, the frequency of this alteration in sporadic PTC and healthy Brazilian individuals indicated that this allele is common in Brazilian individuals. Based on these findings, the *SRGAP1*, *PARP4* and *SRRM2* candidate genes related to familial NMTC were also investigated. Although non-synonymous mutations have been found, none of them were identical to those previously reported.

Case 1 presented 14 variants (*SRGAP1*: p.L953V; *SRRM2*: p.P804T, p.R1934H; *PARP4*: p.I1564T, p.L1550P, p.S1459Y, p.S1394A, p.P1328T, p.G1280R, p.R1108C, p.L1080R, p.V1065A, p.M936T, p.A899T), three of them called deleterious in at least one bioinformatics prediction database for pathogenicity (dbNSFP, Supplementary

Table 3). Two of them were classified as rare mutations according to 1000 Genomes [27] and 6500 Exomes [28]: *SRGAP1* p.L953V (MAF 0.016) and *SRRM2* p.R1934H (not described).

The index patient 2 showed 12 non-synonymous mutations (*SRRM2*: p.P804T; *PARP4*: p.I1564T, p.L1550P, p.S1459Y, p.S1394A, p.R1108C, p.L1080R, p.V1065A, p.M936T, p.S873N, p.I81V, p.A899T). Two of them, mapped in *PARP4* gene, were predicted as pathogenic in at least one dbNSFP annotation: p.A899T (MAF: 0.0323) and p.I81V (MAF: 0.6792). However, the p.I81V is considered common in populational public databases [27, 28].

To gain further insight into the function of other putative candidates associated with familial NMTC, we scrutinized WES data. Rare variants reported with MAF ≤ 0.01 in 1000 Genomes phase 3 [27] and 6500 Exomes [28] were prioritized. Variants ranked per pathogenicity

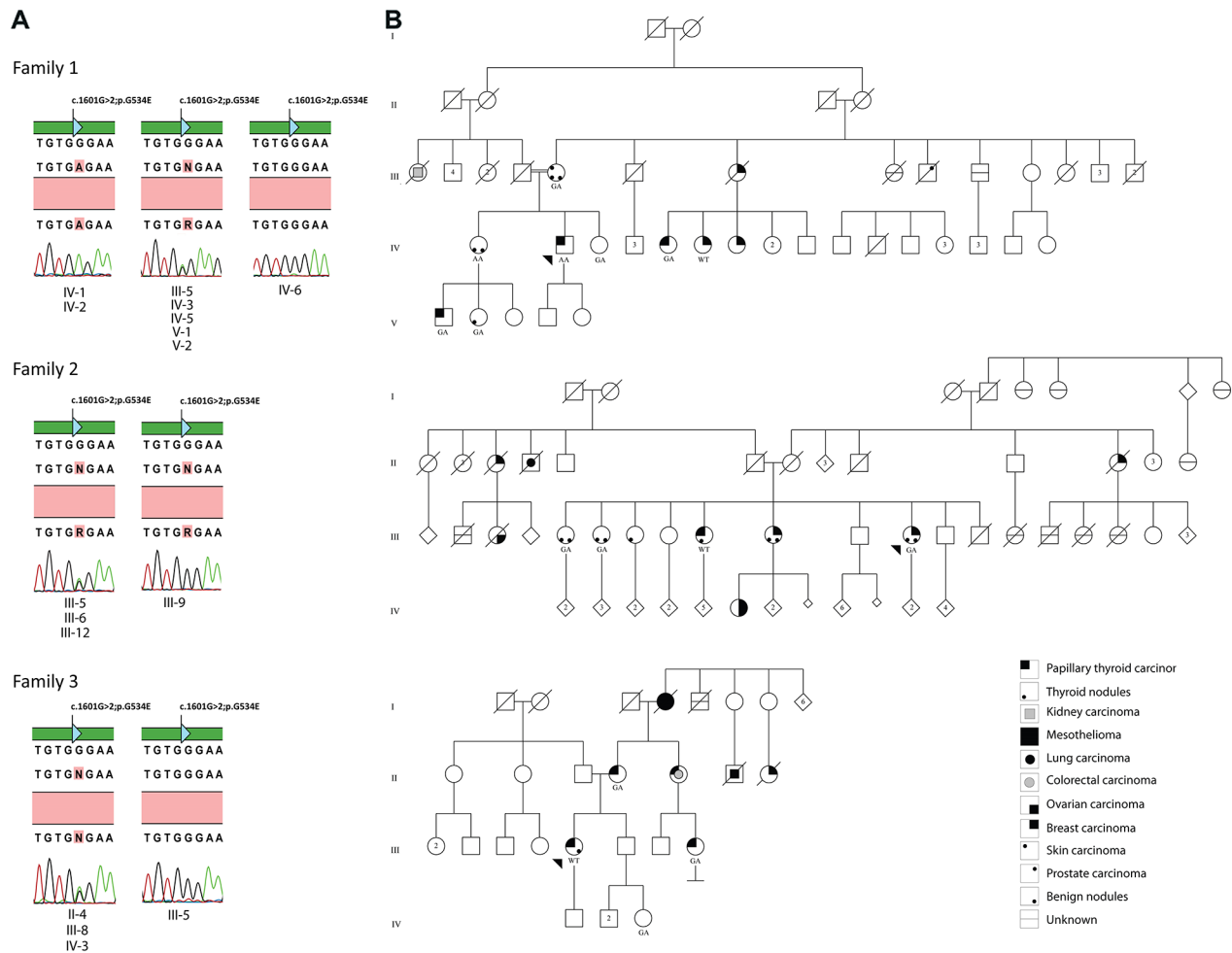


Figure 1: Segregation analysis and *HABP2* p.G354 status in Family 1, 2 and 3. WT. Homozygous for the wild-type allele. GA. Heterozygous. AA. Homozygous for the p.G354E variant. Circles and squares represent female and male members, respectively. The probands are indicated by black arrows. Deceased members are represented by diagonal lines. (A) Electroferograms representing *HABP2* mutation status for each tested individual. (B) Pedigree from families 1, 2 and 3.

prediction, using dbNSFP, are described in details in Supplementary Table 4. A comprehensive pathway enrichment analysis using genes that harbor mutations predicted as deleterious in at least three dbNSFP was performed using KOBAS tool [29] and pathDIP [30]. Extracellular matrix organization pathway was significantly altered for case 1 ($P=5.3 \times 10^{-5}$; $FDR=0.016$). From higher to lower pathogenic degree, the mutated genes in this pathway included *CTSB*, *TNXB*, *COL4A3*, *COL16A1*, *COL24A1*, *COL5A2*, *NID1*, *LOXL2*, *MMP11*, *TRIM24* and *MUSK*, (Figure 2A). Although not enriched in the same pathway, *HABP2* was related to extracellular matrix degradation interacting with *COL4A3*. In addition, cancer related genes [31] as well as cytokines and growth factors, protein kinases, transcription factors and oncogenes presented pathogenic mutations. From these, *MTA2*, *MYH11* and *AMHR2* have highly pathogenic mutations displayed by pathogenic score (Figure 2A).

Similar analysis was performed for the altered genes found in the index Case 2. Although no significantly enriched pathway was deregulated, two tumor suppressor genes, *MSH2* and *NBN* were mutated (Figure 2B). Although predicted as pathogenic, the p.G322D *MSH2* variant (c.965G>A) is classified as benign for the Lynch Syndrome phenotype (ClinVar Variation ID: 1762). Currently, the alteration c.596C>T; p.P199L in *NBN* gene has not been described in public databases. A considerable number of genes harboring pathogenic variants that codify

protein kinases was detected including *ERN2*, *MAP2K3*, *EPHA8*, *STK16*, *GRK7*, *MYLK*, *TTK*, *NEK1*, *ULK2*, *TRIM24* and *TLK2*. Alterations in cell differentiation markers, cytokines and growth factors, transcription factors and oncogenes were also detected (Figure 2B). The protein-protein interaction (PPI) network analysis using IID [32] revealed connections among *FOXH1*, *LHX3*, *NBN*, *NEK1*, *STK16*, which could be relevant in the phenotype (Figure 2B).

A comprehensive pathway enrichment analysis was performed using all proteins from Figure 2. The combination of the two cases identified their physical protein interactions using IID [32] (Figure 2; Supplementary Table 5). The resulting proteins were used to query pathDIP [30] that identified 31 significantly enriched pathways ($p < 0.043$, FDR : BH-method; Supplementary Table 5), including leptin ($p < 0.0059$), integrins ($p < 0.041$), ECM-receptor interaction ($p < 0.036$), EGFR and ERBB4 ($p < 0.043$). Importantly, the most enriched terms in these pathways included focal-adhesion, EGF, Estrogen, PI3K, and prolactin (Supplementary Table 5).

DISCUSSION

Two probands with personal and or family history of BC and PTC evaluated by exome sequencing analysis presented *HABP2* p.G534E variant. The probands and

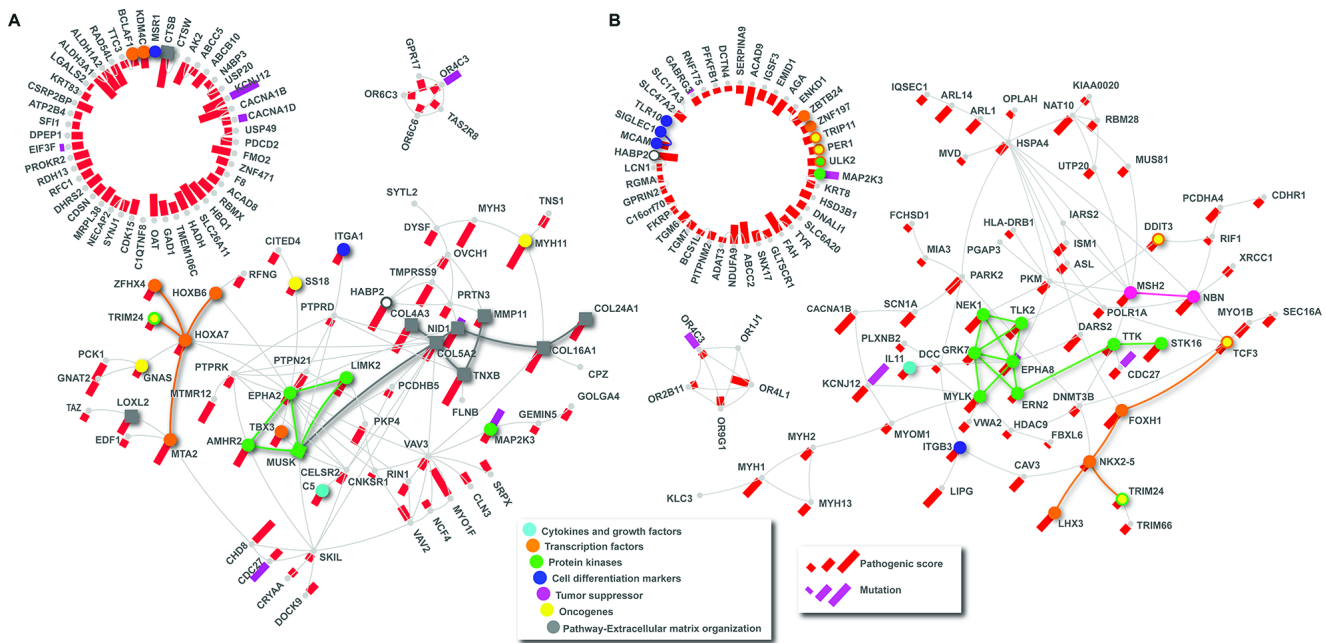


Figure 2: Protein-protein interaction (PPI) network analysis. PPI networks based on mutated genes detected by WES analysis, which were classified as deleterious in at least three dbNSFP pathogenicity predictions, and their interactions partners identified in IID and visualized with NAViGaTOR. Highlighted diamonds represent mutated genes classified according to MSigDB. Squares represent genes enriched for Extracellular Matrix Organization pathway. Pathogenic score and mutation is highlighted by spokes as per legend. (A) Mutated genes identified in the proband 1. (B) Mutated genes identified in the index patient 2.

several relatives from these two families diagnosed with PTC, thyroid goiter, thyroid nodules, BC and other tumors were tested using Sanger sequencing method. Our data revealed the absence of complete co-segregation between the mutation and the phenotype. Two individuals from Family 1, including one diagnosed with PTC and his healthy sister, harbored homozygous p.G534E mutation as result of consanguineous marriage. The homozygous cases have no clinical or pathological differences compared with heterozygous cases. The third family evaluated only by Sanger sequencing revealed three of four individuals with the p.G534E variant; however, the index patient who had PTC was negative for the mutation. Furthermore, two healthy individuals from families 1 and 3 (both, IV-3) were positive for the mutation. Overall, these findings give additional support to the absence of the association between NMTC and *HABP2* p.G534E variant.

Interestingly, we have identified patients diagnosed with BC positive for the *HABP2* p.G534E mutation, which is additional evidence that this variant is not associated with thyroid disease. To our knowledge, breast cancer has not been described as associated with p.G534E mutation.

To investigate the *HABP2* p.G534E allele frequency in Brazilian population, 170 healthy individuals were evaluated and five of them were heterozygous for the mutation (AF = 0.0147). The ethnicity of Brazilian individuals is questionable mostly by the high level of miscegenation; however, the Brazilian genetic background also derives from the European population [33, 34]. According to the last update of 1000 Genomes project Phase 3 [27] and considering European allele frequency, the minor allele frequency of p.G534E is 0.027, which is not significantly different from the Brazilian individuals tested in our study (Fisher's exact test, $P = 0.3$). Similarly, no significant results were found in the comparison with the European allele frequency in ExAC database (unfinished) [35], in which p.G534E allele frequency is 0.033 (Fisher's exact test, $P = 0.065$).

The *HABP2* mutation allele frequency was 0.04 in 50 sporadic PTC samples, which is consistent with data reported by TCGA (0.047). The comparison between healthy Brazilian individuals and sporadic PTC cases revealed no significant difference (Fisher's exact test, $P = 0.124$). Based on these findings, Brazilian allele frequency is concordant with public databases and no association of p.G534E with sporadic PTC was found.

Since Gara et al. (2015) reported the *HABP2* p.G534E as a candidate to hereditary NMTC [15], several studies have been reported aiming to establish the involvement of this mutation as a non random event in different populations. Except for Zhang et al. (2016) [16], these studies have failed to prove the association between NMTC and *HABP2* variant, including our study. Data reported in Spanish population [22, 24], highly inbred Saudi Arabian population [21], United Kingdom and Ireland [23], as well as patients followed by Ohio

State University, United States [17], have demonstrated no involvement of p.G534E as a predisposition gene in thyroid tumors.

In the past few years, new variants associated with familial NMTC have been reported including *SRGAP1* (Q149H and R617C), *SRRM2* (p.S346F) and *PARP4* (p.G496V and p.T1170I) [12-14]. These specific variants were not detected in our probands (cases 1 and 2). However, the rare mutation *SRRM2* p.R1934H (pathogenic in four pathogenicity predictions and no described in dbSNP so far) was detected in the index Case 1 (Supplementary Table 3). Interestingly, the allele frequency of this variant was 2.594×10^{-4} and 2.999×10^{-5} in Latino and European populations (ExAC project, not finished) [35], respectively. Further studies are necessary to better evaluate the involvement of the *SRRM2* in familial NMTC.

A considerable number of alterations with potential to modify the effect of tumor suppressor genes have been described [36]. In addition, the combination of different mutations in cancer predisposition genes has implicated in phenotype overlapping and a synergic interaction, where several loci may contribute with the phenotype [37]. For instance, a combination of papillary thyroid carcinoma and paraganglioma in the same patient revealed alterations in *PTEN* and *SDHC* genes, leading to Cowden syndrome and pheochromocytoma–paraganglioma diagnosis [38]. A recent study suggested that the mutation background of an individual might modulate the severity of the Mendelian diseases [39]. In our study, no classical mutations associated with hereditary cancer syndromes were identified, however several rare and predicted as pathogenic mutations were detected in the Cases 1 and 2. For instance, Case 1 harbors pathogenic mutations in genes involved in extracellular matrix organization pathway, similar with *HABP2* protein function [40]. The extracellular matrix proteins are involved in angiogenesis, cell growth, fibrosis, cell signaling and disrupting of tissue morphogenesis [41]. In thyroid cancer, dysfunction of *COLL1* and *LOX* genes, promoted by *BRAF* and *PTEN* knockdown, were associated with ripping stiffen of Col1 matrix, enhancing cell motility and tumor progression [42]. In our analysis, *CTSB* gene harbors a mutation predicted as highly pathogenic (Figure 2A). Protein encoded by *CTSB* regulates negatively Wnt/ β -catenin signaling pathway [43] related with tumor development [44]. Extracellular matrix is essential for breast normal function (as mammary gland remodeling before and after lactation) and tumor development, progression and metastasis, as its composition differs between metastatic and non-metastatic breast tumor stages [45]. Inhibitors of extracellular components as well as integrins, proteoglycans, metalloproteinase, collagens and C1-peptidase protein families have demonstrated promising pre-clinical results in breast cancer [45]. For instance, cathepsin B (*CTSB* gene, reported as mutated in our study)

inhibition has been reported as associated with reduced bone metastasis in breast cancer from mice [46].

Although predicted pathogenic alterations were found involving *MSH2* and *NBN*, no significantly enriched pathway was found for index Case 2. *MSH2* p.G322D variant is described as benign in the Lynch Syndrome diagnosis (ClinVar Variation ID: 1762) [47]. Recently, the *MSH2* p.G322D variant was associated with breast cancer risk in Polish women [48]. *NBN* is DNA repair family gene, involved in the *MRE11/RAD50/NBN* complex, with nine variants (rs1805794, rs1805790, rs36226237, rs924, rs376639, rs1061302, rs1063054, rs2735383, rs805794) described as associated with high cancer risk [49]. Recently, Kraus et al. investigate the mutational status of 581 BC patients using a panel-based screening of 14 breast and ovary susceptibility genes, including *NBN*. One patient presented a deleterious variant c.657_661del5 of *NBN* [50]. In addition, the c.657del5 variant in the *NBN* gene was reported as associated with pancreatic cancer predisposition [51]. Differently of these reports, we found the variant c.596C>T p.P199L in our case 2. Eleven mutated genes classified as protein kinases were also identified in our Case 2 (Figure 2B). Interestingly, a group of proteins highly connected and with the same function might lead to cellular dysfunction and pre-cancer stages. In our study, the index patient 2 was diagnosed with BC, benign breast nodule, thyroid nodules and colorectal polyps. The presence of mutations in *MSH2* and *NBN* has the potential to increase the risk of the patient developing multiple primary tumors.

A comprehensive pathway enrichment analysis using pathDIP and proteins from the network (Figure 2) revealed 31 significantly enriched pathways ($p < 0.043$), including leptin, integrins, interleukins, ECM-receptor interaction, EGFR and ERBB4. Interestingly, the most enriched terms in the titles of these pathways included focal-adhesion, EGF, Estrogen, PI3K, and prolactin (Supplementary Table 5). In the current study, we focused on comprehensive pathway enrichment analysis. In the future, using additional patient cases and larger PPI network, we could use additional tools that combine mutation data (such as HotNet [52]) or arbitrary omics data (e.g., KeyPathwayMiner [53], NBS [54], MUFFIN [55], VarWalker [56]) with signaling networks.

In conclusion, no co-segregation of the *HABP2* p.G534E was found in three families with NMTC and breast cancer. The variant is not significantly associated with PTC in Brazilian population and showed similar allele frequency in tumor samples and healthy individuals. Our data also refute the dominant negative effect of this variant as proposed by Gara et al. (2015) [15]. Alterations reported here may contribute to high risk of NMTC. A limitation of our study was the impossibility to evaluate the exome of all affected family members, compromising the identification of other putative *locus* of predisposition

that segregate with the phenotype. Our results point out that the *HABP2* p.G534E mutation is not associated with hereditary NMTC. Genetic mechanisms involved in familial NMTC remain elusive and further investigation is necessary.

MATERIALS AND METHODS

Ethics statement

The use of the clinical samples was approved by the Institutional Human Research Ethics Committees at A.C. Camargo Cancer Center, Sao Paulo and Faculty of Medicine, UNESP, Botucatu, Sao Paulo, Brazil (FMB-PC-197/2012; CEP1175/08ext). Informed consent was obtained from all patients, and the experiments were performed in accordance with relevant guidelines and regulations.

Patients and clinical samples

Twenty index cases with personal and family history of NMTC, thyroid disease as well as breast carcinoma were evaluated by whole exome sequencing. We focused this study on two patients that presented *HABP2* p.G534E variant, one diagnosed with invasive ductal breast carcinoma and the other with PTC. Family members from index patient 1 (7 relatives) and 2 (three relatives) were included in this study. Although these patients presented tumors frequently described in the Cowden Syndrome (CS) spectrum, none of them fulfilled the clinical criteria for CS. In addition, the index patients were negative for pathogenic mutations in *PTEN*, *BRAC1*, *BRCA2* and *TP53* genes. A third index patient diagnosed with PTC, with family cancer history of PTC and BC and three relatives, were also investigated for *HABP2* status using Sanger sequencing. All patients were prospectively followed by the Department of Oncogenetics, A.C. Camargo Cancer Center, Sao Paulo, Brazil. Clinical features and family history of all individuals are presented on Supplementary Table 1.

Index Case 1 was diagnosed with PTC at 41-years-old. His familial history included a nephew diagnosed with PTC (27 years old), an aunt diagnosed with BC (90 years old) and skin carcinoma (87 years old) who had three daughters, one diagnosed with PTC (55 years old), and two with BC (at age 60 and 61, respectively) (Figure 1B, Supplementary Table 1). The proband's mother developed a skin tumor (75 years old) and thyroid colloid goiter (67 years old). She was married with a first-degree cousin.

Index Case 2 presented a BC (46 years old) and a history of several benign lesions (benign nodules from breast, skin, thyroid and colorectal polyp). Her sisters presented history of cancer: one with papillary thyroid

carcinoma (66 years old) and the other with BC (58 years old). Her niece was diagnosed with breast and ovary carcinomas (47 years-old); two aunts with BC, beyond other malignancies in the family (Figure 1B).

The proband from the third family was diagnosed with PTC (40 years old) concurrent with breast benign nodules. Her family history encompasses a mother with PTC (40 years old), a female cousin with PTC (38 years old), an aunt with colorectal cancer (40 years old) and PTC (50 years old), and a second-degree female cousin diagnosed with BC (50 years old). This family also presented individuals with throat, pleura and other tumors types in the maternal lineage (Figure 1B, Supplementary Table 1).

A cohort of 170 healthy individuals (84 males and 86 females) was recruited at the same hospital and screened for the variant p.G534E. These individuals were cancer-free so far, with no known thyroid diseases or family history of cancer. The median age of this set of cases was 49 years (range 21-87 years old). Fifty sporadic PTC (11 male and 39 female) with median age of 40 years (range 21-75 years old) were also assessed. These patients have no family history of cancer. Clinical and histological characteristics of the sporadic PTC are detailed in Supplementary Table 2.

Genomic DNA extraction

Genomic DNA was extracted from blood leukocytes using Qiacube DNA Blood kit (Qiagen, Valencia, CA) for two index patients and all healthy individuals. Family members from *HABP2* p.G534E positive patients, as well as the individuals for family 3, had their DNA extracted from saliva using Oragene-DNA (DNA Genotek, Ottawa, CA). Gentra Puregene Tissue Kit (Qiagen, Valencia, CA) was used to obtain DNA from fresh frozen PTC tissue samples. The DNA extraction followed the manufacturer's recommendations.

Whole exome sequencing (WES) analysis

Whole exome libraries were prepared with Exome Nextera Enrichment kit-Illumina (San Diego, California, USA) with 62 Mb target regions. Paired-end sequencing of 100 bp was carried out in HiSeq2500 Illumina (San Diego, California, USA), at the Center for Functional Genomics core facility, ESALQ-University of Sao Paulo, Piracicaba, Brazil.

Raw reads data were trimmed with Seqclean software [57]. Paired-end reads were aligned to UCSC hg19 using Botiwe2 [58], in local-alignment mode. WES analysis achieved coverage of 53x. Duplicate reads were marked using the MarkDuplicates utility from Picard [59] to filter out PCR artifacts. Variants were detected with GATK [60] and SAMtools [61]. ANNOVAR [62] was used for data annotation.

Exonic alterations were prioritized according to their function focusing on frameshift, splice-site and non-synonymous variants. Data were compared with 1000 Genomes phase 3 [27] and 6500 Exomes [28] and variants reported with an allele frequency ≤ 0.01 were selected. After the annotation with dbNSFP database [63], the alterations classified as deleterious in at least three pathogenicity predictions were selected to perform pathway enrichment and network analysis. Pathways enrichment analysis was performed using KOBAS 2.0 tool [29]. Selected variants were then used to identify biological context by querying protein-protein interaction (PPI) database Interologous Interaction Database v. 2016-03 [32] (<http://ophid.utoronto.ca/iid>). The resulting network was visualized in NAViGaTOR v3 [64] (<http://ophid.utoronto.ca/navigator>).

To elucidate biological functions of the protein network (Figure 2), we performed a comprehensive pathway enrichment analysis. The Pathway Data Integration Portal (pathDIP) was used to perform enrichment analysis across pathways from 20 major pathway databases (pathDIP ver. 2.5; [30]). We considered literature curated gene: pathway memberships and those predicted according to experimentally detected protein-protein interactions (including interactions experimentally detected between orthologues plus FpClass [65] interactions with minimum confidence level for predicted associations equal 0.95 (for more details see pathDIP documentation in <http://ophid.utoronto.ca/pathDIP>).

Data confirmation and validation

The presence of the variant p.G534E detected by WES was confirmed by Sanger re-sequencing in the two index cases, their family members and in the family 3 members. For the *HABP2* exon 13 amplification the following primers were constructed: forward primer 3'-CCCTGACACCCCCTGGAGAG-5'; and reverse primer 3'-GCTCTGGAGGTGCCATTGT-5' (Gara et al., 2015). Polymerase chain reaction was performed according to standard protocols using Platinum Taq DNA Polymerase (Life Technologies, Carlsbad, California, USA) with the following conditions: 2 min at 94°C; followed by 35 cycles of 45 sec at 94°C, 30 sec at 63°C and 45sec at 72°C; followed by 10 min at 72°C for final extension. Sanger sequencing was carried out using BigDye terminator v3.1 and ABI Prism 3130XL sequencer, both from Applied Biosystem (Foster City, California, USA).

Abbreviations

PTC – Papillary Thyroid Carcinoma; WES – Whole Exome Sequencing; TC – Thyroid Cancer; NMTC – Non-Medullary Thyroid Carcinoma; FTC – Follicular Thyroid Carcinoma; HCPS – Hereditary Cancer Predisposition Syndromes; BC – Breast Carcinoma; KOBAS – KEGG

Orthology-Based Annotation System; AF – Allele Frequency; MAF – Minor Allele Frequency; dbNSFP – Single Nucleotide Polymorphism Database; PPI – Protein-Protein Interaction; ExAC – Exome Aggregation Consortium; TCGA – The Cancer Genome Atlas; CS – Cowden Syndrome.

Author contributions

S.R.R., M.I.A. and L.P.K conceived and designed the experiments. M.P and R.T performed the experiments. M.P., S.A.A., F.A.M and I.G performed the bioinformatics analysis. S.R.R., S.A.D., M.I.A., I.J. and L.P. contributed with reagents/materials/analysis tools; M.P., S.A.D and S.R.R wrote the paper. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

The authors would like to thank the patients and their relatives who participated of the study and the Nucleic Acid Bank of A.C. Camargo Cancer Center (São Paulo, Brazil) for sample processing. The authors are grateful for the assistance given by Mateus Camargo Barros-Filho and Karina Miranda Santiago.

CONFLICTS OF INTEREST

The authors declare no competing financial interest.

FUNDING

SRR received investigator fellowship awards from CNPq. The study was supported by grants from São Paulo Research Foundation (FAPESP 2013/0186-7 and MP Fellowship 2012/12714-5, 2014/03983-8) and National Council for Scientific and Technological Development (CNPq Universal 481132/2012-0). IJ is in part supported by the Canada Research Chair Program (CRC #203373 and #225404), Natural Sciences Research Council (NSERC #203475), Ontario Research Fund (GL2-01-030), Canada Foundation for Innovation (CFI #225404, #30865) and IBM.

REFERENCES

1. Forman D, Bray F, Brewster DH, Gombe Mbalawa C, Kohler B, Piñeros M, Steliarova-Foucher E, Swaminathan R, Ferlay J. Cancer Incidence in Five Continents Vol. X. IARC Scientific Publication No. 164. 2014.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin.* 2016; 66:7-30.
3. Schneider DF, Chen H. New developments in the diagnosis and treatment of thyroid cancer. *CA Cancer J Clin.* 2013; 63:374-94.
4. Davies L, Welch HG. Current thyroid cancer trends in the United States. *JAMA Otolaryngology-Head & Neck Surgery.* 2014; 140:317-22.
5. Nagy R, Ringel MD. Genetic predisposition for nonmedullary thyroid cancer. *Hormones and Cancer.* 2015; 6:13-20.
6. Durieux E, Descotes F, Mauduit C, Decaussin M, Guyétant S, Devouassoux-Shisheboran M. The co-occurrence of an ovarian Sertoli-Leydig cell tumor with a thyroid carcinoma is highly suggestive of a DICER1 syndrome. *Virchows Arch.* 2016; 468:631-6.
7. Broaddus RR, Lynch PM, Lu KH, Luthra R, Michelson SJ. Unusual tumors associated with the hereditary nonpolyposis colorectal cancer syndrome. *Modern Pathology.* 2004; 17:981-9.
8. Achatz MI, Olivier M, Le Calvez F, Martel-Planche G, Lopes A, Rossi BM, Ashton-Prolla P, Giugliani R, Palmero EI, Vargas FR, Da Rocha JC, Vettore AL, Hainaut P. The TP53 mutation, R337H, is associated with Li-Fraumeni and Li-Fraumeni-like syndromes in Brazilian families. *Cancer Letters.* 2007; 245:96-102.
9. Masciari S, Van den Abbeele AD, Diller LR, Rastarhuyeva I, Yap J, Schneider K, Digianni L, Li FP, Fraumeni JF Jr, Syngal S, Garber JE. F18-fluorodeoxyglucose-positron emission tomography/computed tomography screening in Li-Fraumeni syndrome. *JAMA.* 2008; 299:1315-9.
10. Basso TR, Villacis RA, Canto LM, Alves VM, Lapa RM, Nóbrega AF, Achatz MI, Rogatto SR. Genomic profile of a Li-Fraumeni-like syndrome patient with a 45,X/46,XX karyotype, presenting neither mutations in TP53 nor clinical stigmata of Turner syndrome. *Cancer Genetics.* 2015; 208:341-4.
11. Pelizzo MR, Pennelli G, Zane M, Galuppini F, Colletti PM, Merante Boschini I, Rubello D. Papillary thyroid carcinoma (PTC) in Lynch syndrome: Report of two cases and discussion on Lynch syndrome behaviour and genetics. *Biomedicine & Pharmacotherapy.* 2015, 74:9-16.
12. He H, Li W, Wu D, Nagy R, Liyanarachchi S, Akagi K, Jendrzewski J, Jiao H, Hoag K, Wen B, Srinivas M, Waidyaratne G, Wang R, et al. Ultra-rare mutation in long-range enhancer predisposes to thyroid carcinoma with high penetrance. *PLoS One.* 2013; 8:e61920. Erratum in: *PLoS One.* 2013; 8.
13. Ikeda Y, Kiyotani K, Yew PY, Kato T, Tamura K, Yap KL, Nielsen SM, Mester JL, Eng C, Nakamura Y, Grogan RH. Germline PARP4 mutations in patients with primary thyroid and breast cancers. *Endocrine-Related Cancer.* 2016; 23:171-9.
14. Tomsic J, Fultz R, Liyanarachchi S, He H, Senter L, de la Chapelle A. HAP2 G534E Variant in papillary thyroid carcinoma. *PLoS One.* 2016; 11:e0146315.
15. Gara SK, Jia L, Merino MJ, Agarwal SK, Zhang L, Cam M, Patel D, Kebebew E. Germline HAP2 mutation causing familial nonmedullary thyroid cancer. *The New England Journal of Medicine.* 2015; 373:448-55.

16. Zhang T, Xing M. HBP2 G534E Mutation in familial nonmedullary thyroid cancer. *The Journal of the National Cancer Institute*. 2016; 108:djv415.
17. Tomsic J, He H, Akagi K, Liyanarachchi S, Pan Q, Bertani B, Nagy R, Symer DE, Blencowe BJ, de la Chapelle A. A germline mutation in SRRM2, a splicing factor gene, is implicated in papillary thyroid carcinoma predisposition. *Scientific Reports*. 2015; 5:10566.
18. Sponziello M, Durante C, Filetti S. HBP2 mutation and nonmedullary thyroid cancer. *The New England Journal of Medicine*. 2015; 373:2085-6.
19. Tomsic J, He H, de la Chapelle A. HBP2 mutation and nonmedullary thyroid cancer. *The New England Journal of Medicine*. 2015; 373:2086.
20. Zhou EY, Lin Z, Yang Y. HBP2 mutation and nonmedullary thyroid cancer. *The New England Journal of Medicine*. 2015; 373:2084-5.
21. Alzahrani AS, Murugan AK, Qasem E, Al-Hindi H. HBP2 gene mutations do not cause familial or sporadic non-medullary thyroid cancer in a highly inbred middle eastern population. *Thyroid*. 2016; 26:667-71.
22. Bohórquez ME, Estrada AP, Stultz J, Sahasrabudhe R, Williamson J, Lott P, Duque CS, Donado J, Mateus G, Bolaños F, Vélez A, Echeverry M, Carvajal-Carmona LG. The HBP2 G534E polymorphism does not increase nonmedullary thyroid cancer risk in Hispanics. *Endocrine Connections*. 2016; 5:123-7.
23. Sahasrabudhe R, Stultz J, Williamson J, Lott P, Estrada A, Bohorquez M, Palles C, Polanco-Echeverry G, Jaeger E, Martin L, Magdalena Echeverry M, Tomlinson I, Carvajal-Carmona LG, et al. The HBP2 G534E variant is an unlikely cause of familial non-medullary thyroid cancer. *The Journal of Clinical Endocrinology & Metabolism*. 2015; 21:jc20153928.
24. Ruiz-Ferrer M, Fernández RM, Navarro E, Antiñolo G, Borrego S. G534E Variant in HBP2 and nonmedullary thyroid cancer. *Thyroid*. 2016; 26:987-8.
25. Weeks AL, Wilson SG, Ward L, Goldblatt J, Hui J, Walsh JP. HBP2 germline variants are uncommon in familial nonmedullary thyroid cancer. *BMC Medical Genetics*. 2016; 17:60.
26. Carvajal-Carmona LG, Tomlinson I, Sahasrabudhe R. RE: HBP2 G534E Mutation in Familial Nonmedullary Thyroid Cancer. *Journal of the National Cancer Institute*. 2016; 108.
27. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. A global reference for human genetic variation. *Nature*. 2015; 526:68-74.
28. Fu W, O'Connor TD, Jun G, Kang HM, Abecasis G, Leal SM, Gabriel S, Rieder MJ, Altshuler D, Shendure J, Nickerson DA, Bamshad MJ, NHLBI Exome Sequencing Project, et al. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature*. 2013; 493:216-20. Erratum in: *Nature*. 2013; 495:270. Rieder, Mark J [added].
29. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li CY, Wei L. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Research*. 2011; 39:W316-22.
30. Rahmati S, Abovsky M, Pastrello C, Jurisica I. pathDIP: An annotated resource for known and predicted human gene-pathway associations and pathway enrichment analysis. *Nucleic Acids Research*. 2017; 45:D419-D426.
31. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*. 2005; 102:15545-50.
32. Kotlyar M, Pastrello C, Sheahan N, Jurisica I. Integrated interactions database: tissue-specific view of the human and model organism interactomes. *Nucleic Acids Research*. 2016; 44:D536-41.
33. Santos RV, Fry PH, Monteiro S, Maio MC, Rodrigues JC, Bastos-Rodrigues L, Pena SD. Color, race, and genomic ancestry in Brazil: dialogues between anthropology and genetics. *Current Anthropology*. 2009 50:787-819.
34. Palha T, Gusmão L, Ribeiro-Rodrigues E, Guerreiro JF, Ribeiro-Dos-Santos A, Santos S. Disclosing the genetic structure of Brazil through analysis of male lineages with highly discriminating haplotypes. *PLoS One*. 2012; 7:e40007.
35. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016; 536:285-91.
36. Lord CJ, Ashworth A. BRCAness revisited. *Nature Reviews Cancer*. 2016; 16:110-20.
37. Whitworth J, Skytte AB, Sunde L, Lim DH, Arends MJ, Happerfield L, Frayling IM, van Minkelen R, Woodward ER, Tischkowitz MD, Maher ER. Multilocus Inherited Neoplasia Alleles Syndrome: A Case Series and Review. *JAMA Oncology*. 2016; 2:373-9.
38. Zbuk KM, Patocs A, Shealy A, Sylvester H, Miesfeldt S, Eng C. Germline mutations in PTEN and SDHC in a woman with epithelial thyroid cancer and carotid paraganglioma. *Nature Reviews Clinical Oncology*. 2007; 4:608-12.
39. Vu V, Verster AJ, Schertzberg M, Chuluunbaatar T, Spensley M, Pajkic D, Hart GT, Moffat J, Fraser AG. Natural Variation in Gene Expression Modulates the Severity of Mutant Phenotypes. *Cell*. 2015; 162:391-402.
40. Etscheid M, Beer N, Dodt J. The hyaluronan-binding protease upregulates ERK1/2 and PI3K/Akt signalling pathways in fibroblasts and stimulates cell proliferation and migration. *Cell Signal*. 2005; 17:1486-94.

41. Theocharis AD, Skandalis SS, Gialeli C, Karamanos NK. Extracellular matrix structure. *Advanced Drug Delivery Reviews*. 2016; 97:4-27.
42. Jolly LA, Novitskiy S, Owens P, Massoll N, Cheng N, Fang W, Moses HL, Franco AT. Fibroblast-Mediated Collagen Remodeling Within the Tumor Microenvironment Facilitates Progression of Thyroid Cancers Driven by BrafV600E and Pten Loss. *Cancer Research*. 2016; 76:1804-13.
43. Zhang ZY, Mai Y, Yang H, Dong PY, Zheng XL, Yang GS. CTSB promotes porcine preadipocytes differentiation by degrading fibronectin and attenuating the Wnt/ β -catenin signaling pathway. *Molecular and Cellular Biochemistry*. 2014; 395:53-64.
44. Xue G, Romano E, Massi D, Mandalà M. Wnt/ β -catenin signaling in melanoma: Preclinical rationale and novel therapeutic insights. *Cancer Treatment Reviews*. 2016; 49:1-12.
45. Insua-Rodríguez J, Oskarsson T. The extracellular matrix in breast cancer. *Advanced Drug Delivery Reviews*. 2016; 97:41-55.
46. Withana NP, Blum G, Sameni M, Slaney C, Anbalagan A, Olive MB, Bidwell BN, Edgington L, Wang L, Moin K, Sloane BF, Anderson RL, Bogyo MS, et al. Cathepsin B inhibition limits bone metastasis in breast cancer. *Cancer Research*. 2012; 72:1199-209.
47. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J, Jang W, Katz K, Ovetsky M, et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Research*. 2016; 44:D862-8.
48. Smolarz B, Makowska M, Samulak D, Michalska MM, Romanowicz H. Gly322Asp and Asn127Ser single nucleotide polymorphisms (SNPs) of hMSH2 mismatch repair gene and the risk of triple-negative breast cancer in Polish women. *Familial Cancer*. 2015; 14:81-8.
49. Berardinelli F, di Masi A, Antoccia A. NBN gene polymorphisms and cancer susceptibility: A systemic review. *Current Genomics*. 2013; 14:425-40.
50. Kraus C, Hoyer J, Vasileiou G, Wunderle M, Lux MP, Fasching PA, Krumbiegel M, Uebe S, Reuter M, Beckmann MW, Reis A. Gene panel sequencing in familial breast/ovarian cancer patients identifies multiple novel mutations also in genes others than BRCA1/2. *International Journal of Cancer*. 2017; 140:95-102.
51. Borecka M, Zemankova P, Lhota F, Soukupova J, Kleiblova P, Vocka M, Soucek P, Ticha I, Kleibl Z, Janatova M. The c.657del5 variant in the NBN gene predisposes to pancreatic cancer. *Gene*. 2016; 587:169-72.
52. Vandin F, Upfal E, Raphael BJ. Algorithms for detecting significantly mutated pathways in cancer. *Journal of Computational Biology*. 2011; 18:507-22.
53. Alcaraz N, Kucük H, Weile J, Wipat A, Baumbach J. KeyPathwayMiner: Detecting case-specific biological pathways using expression data. *Internet Mathematics*. 2011; 7:299-313.
54. Hofree M, Shen JP, Carter H, Gross A, Ideker T. Network-based stratification of tumor mutations. *Nature Methods*. 2013; 10:1108-1115.
55. Cho A, Shim JE, Kim E, Supek F, Lehner B. MUFFINN: cancer gene discovery via network analysis of somatic mutation data. *Genome Biology*. 2016; 17:129.
56. Jia P, Zhao Z. VarWalker: Personalized mutation network analysis of putative cancer genes from next-generation sequencing data. *PLOS Computational Biology*. 2014; 10:e1003460.
57. Zhbannikov I. Seqclean software <https://github.com/ibest/seqclean> (2015).
58. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature Methods*. 2012; 9:357-9.
59. Picard software <http://broadinstitute.github.io/picard> (2015).
60. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky 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 Research*. 2010; 20:1297-303.
61. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009; 25:2078-9.
62. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*. 2010; 38:e164.
63. Liu X, Jian X, Boerwinkle E. dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. *Human Mutation*. 2013; 34:E2393-402.
64. Brown KR, Jurisica I. Unequal evolutionary conservation of human protein interactions in interologous networks. *Genome Biology*. 2007; 8:R95.
65. Kotlyar M, Pastrello C, Pivetta, F, Lo Sardo A, Cumbaa C, Li H, Naranian T, Niu Y, Ding Z, Vafae F, Broackes-Carter F, Stagljar I, Jurisicova A, et al. Comprehensive *in silico* prediction of physical protein interactions and characterization of interactome orphans. *Nature Methods*. 2015; 12:79-84.