ISSN 2472-1972

Comprehensive Genomic Characterization of Parathyroid Cancer Identifies Novel Candidate Driver Mutations and Core Pathways

Callisia N. Clarke,[1]* Panagiotis Katsonis,[2]* Teng-Kuei Hsu,[3] Amanda M. Koire,[4] Angelica Silva-Figueroa,[5] Ioannis Christakis,[5] Michelle D. Williams,[6] Merve Kutahyalioglu,[7] Lily Kwatampora,[7] Yuanxin Xi,[8] Jeffrey E. Lee,[5] E. Scott Koptez,[9] Naifa L. Busaidy,[7] Nancy D. Perrier,[5]* and Olivier Lichtarge[2]*

[1]Division of Surgical Oncology, Medical College of Wisconsin, Milwaukee, Wisconsin 53226; [2]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030; [3]Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas 77030; [4]Program in Quantitative and Computational Biosciences, Baylor College of Medicine, Houston, Texas 77030; [5]Department of Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030; [6]Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030; [7]Department of Endocrine Neoplasia and Hormonal Disorders, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030; [8]Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030; and [9]Department of Gastrointestinal Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030

ORCiD numbers: 0000-0002-5229-961X (C. N. Clarke).

*C.N.C. and P.K. are co-first authors. N.D.P. and O.L. are co-last authors.

Context: Elucidating the genomic landscape of sporadic parathyroid carcinoma (PC) has been limited by low tumor incidence.

Objective: Identify driver mutations of sporadic PC and potential actionable pathways.

Methods: Patients undergoing surgical resection for sporadic PC between 1980 and 2016 at MD Anderson Cancer Center were identified. Patients with sporadic PC according to World Health Organization diagnostic criteria and with available formalin-fixed, paraffin-embedded (FFPE) PC tumor tissue were included and their clinical data analyzed to assess extent of disease. Patients with parathyroid tumors of uncertain malignancy or atypical parathyroid neoplasms were excluded. Thirty-one patients meeting diagnostic criteria had available tissue for analysis. FFPE PC tumors were subjected to DNA extraction and next-generation whole-exome sequencing. All variant calls are single-algorithm only. Twenty-nine samples passed quality assurance after DNA extraction.

Main Outcome Measures: Somatic or private germline mutations present in sporadic PC and identification of pathways involved in tumorigenesis.

Results: We identified 35 genes with considerable mutational load; only eight genes were previously identified in other PC cohorts. These genes mediate critical processes, including chromosome organization, DNA repair, and cell cycle regulations. Gene mutations involved in MAPK signaling and immune response are also heavily implicated. These findings are limited by inherent molecular artifacts in FFPE tissue analysis and the absence of matched germline DNA. Additionally, variant calls are only single algorithm and may include false-positive/negative calls.

Abbreviations: AF, allelic fraction; ExAC, Exome Aggregation Consortium; FFPE, formalin-fixed, paraffin-embedded; IQR, interquartile range; K-S, Kolmogorov-Smirnov; M, missense; MDACC, MD Anderson Cancer Center; mTOR, mammalian target of rapamycin; N, nonsense; NGS, next-generation sequencing; PC, parathyroid carcinoma; S, silent; TCGA, The Cancer Genome Atlas; TCGC, The Cancer Gene Census; WHO, World Health Organization.

Received 7 February 2018
Accepted 29 August 2018
First Published Online 3 September 2018

Conclusion: We identified 33 candidate driver genes of sporadic PC, in addition to previously known driver genes CDC73 and MEN1.

Freeform/Key Words: genome, mutation, parathyroid carcinoma, whole exome sequencing

Parathyroid carcinoma (PC) is a rare but lethal endocrine malignancy and an uncommon cause of primary hyperparathyroidism that may lead to deleterious and sometimes fatal hypercalcemia. Surgery is the mainstay of treatment, but achieving a cure is rare, as 50% to 60% of patients with PC will develop recurrent disease even after margin negative resection. There are no known effective chemotherapeutic agents against PC, so patients with persistent disease have limited therapeutic options aimed only at palliating hypercalcemia. Overall survival remains low at 85% and 49% over 5 and 10 years, respectively [1, 2]. For this reason, there remains a pressing need to identify molecular drivers and potentially actionable pathways in PC carcinogenesis; however, these efforts have been hindered by low disease incidence and wide variation in pathologic criteria for PC diagnosis [3, 4].

Next-generation sequencing (NGS) has emerged as a powerful tool in cancer genomics, leading to a rapid increase in our knowledge of the genetic and epigenetic events involved in tumor formation and progression and identifying actionable mutations. The Cancer Genome Atlas (TCGA) has identified major genetic drivers of a number of human tumors using comprehensive genomic characterization and identified potential targets of therapeutic interest [5]. Due to its low incidence, PCs were not evaluated in these studies, so the genomic landscape of PC is still being defined. The Cancer Gene Census (TCGC) [6], a manually curated list of genes that are somatically mutated and causally implicated in human cancer by at least two independent reports, currently associates only two genes with PC: CDC73 and MEN1. These studies primarily involved cases of familial hyperparathyroidism-jaw tumor syndrome and familial isolated primary hyperparathyroidism and reported CDC73 and MEN1 as potential drivers of PC tumorigenesis [7–10]. Subsequent cohort studies of sporadic PC also implicated CDC73 as genomic driver of PC tumorigenesis [11–14]. PRUNE2, MTOR, and PIK3CA genes were also found to be mutated in PC in these studies, implicating aberrations in the phosphatidylinositol 3-kinase/AKT/mammalian target of rapamycin (mTOR) as a potential candidate pathway for PC tumor initiation or progression [14–16]. However, more studies are needed to investigate the role of these and other genes and pathways in PC tumorigenesis.

Fresh or fresh-frozen tumor tissue is the preferred specimen for NGS, as this storage method minimizes exposures that compromise DNA integrity. However, access to fresh or fresh-frozen tissue specimens is limited due to storage logistics, and in rare tumors such as PC, formalin-fixed, paraffin-embedded (FFPE) tissue samples are an invaluable biobank and offer considerable opportunity for further genomic exploration of this disease. Concerns regarding DNA fragmentation and modification from formalin fixation are well documented and prove a substantial challenge in the widespread use of FFPE samples for high-throughput molecular characterization. However, the TCGA and others have recognized the unparalleled value of this tumor source, and recent studies demonstrating feasibility of reliable DNA sequencing using stored FFPE samples have generated uniform protocols for improved accuracy and spurred interest in their utilization [17, 18]. In this study, we present genomic characterization in 29 patients with sporadic PC using whole-exome sequencing of FFPE-extracted DNA.

1. Methods

A. Study Population and Data Collection

We performed a retrospective analysis of all patients undergoing surgical resection at MD Anderson Cancer Center (MDACC) for a diagnosis of PC according to World Health

Organization (WHO) diagnostic criteria [15] between 1990 and 2016. Clinical data, banked tumor, and blood samples were collected under institutional review board–approved protocols. An independent pathologist reviewed all slides to further strengthen clinicopathologic criteria for PC diagnosis. Patients with evidence of vascular invasion, local invasion, and distant metastasis were included in our analysis. Patients with parathyroid tumors of uncertain malignancy or atypical parathyroid neoplasms were excluded. Thirty-one patients who met stringent PC diagnostic criteria had available FFPE tumor tissue from primary, recurrent, or metastatic cancers constitute our study cohort. For each patient, three 10-μm sections of tumor were acquired. Primary PC was defined as PC by WHO diagnostic criteria without tumor recurrence after resection, recurrent PC was defined as local tumor recurrence after resection with curative intent, and metastatic PC was defined by pathologic or imaging confirmation of distant metastasis.

B. DNA Extraction

DNA extraction was performed by the Biospecimen Extraction Facility at MDACC. FFPE tumor tissue from primary, recurrent, or metastatic PC was subjected to DNA extraction optimized for FPPE tissues using the Qiagen QIAamp DNA FFPE Tissue kit. DNA was quantified using a NanoDrop spectrophotometer (OD 260 and 280) and Qubit™ dsDNA HS Assay (Thermo Fisher Scientific) with Qubit 2.0 fluorometer. Sample quality assessment was performed using the Fragment Analyzer High Sensitivity Genomic DNA Analysis Kit (Advanced Analytical). The ratio of single-nucleotide substitutions of the same type, C>T (transitions), to single-nucleotide substitutions of a different nucleotide, A>T (transversions), was calculated to measure the effect of C>T artifact in each sample.

C. Whole-Exome Sequencing

Whole-exome sequencing was performed by the Sequencing and Microarray Facility at MDACC using previously published methodology [19]. Briefly, libraries were prepared from Biorupter ultrasonicator (Diagenode)–sheared genomic DNA using the Agilent Technologies SureSelectXT Reagent Kit. Libraries were prepared for capture with 10 cycles of PCR amplification and then assessed for size distribution on the Fragment Analyzer using the High Sensitivity NGS Fragment Analysis Kit (Advanced Analytical) and quantity using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific). Exon target capture was performed using the Agilent Technologies SureSelectXT Human All Exon V4 kit. Following capture, index tags were added to the exon-enriched libraries using 10 cycles of PCR. The indexed libraries were then assessed for size distribution and quantified using the Agilent Technologies 4200TapesStation and the Qubit dsDNA HS Assay Kit, respectively. Equal molar concentrations of libraries were multiplexed eight to nine samples per pool, and each pool was sequenced in one lane of the Illumina HiSeq4000 sequencer, using the 76nt paired end format.

D. Identification of Somatic and Germline Variants

Sequence reads were aligned to human reference genome GRCh37 using BWA (version 0.7.9a). Variants were detected using VarScan v2.3.9 using the VARSCAN2 “mpileup2snp” function with the following parameters for variant calling and filtering: min-coverage 10, *P* value 0.01, min-freq-for-hom 0.9., and strategies for variant filtering to minimize false positives associated with common sequencing artifacts as previously described [20]. Variants were filtered based on exome capture regions, strand bias, *P* value, and coverage and annotated by ANNOVAR (version: Revision 521; human reference: *hg19*) [21]. As the matched normal samples were not available, a pooled virtual normal sample was built as described by Hiltemann *et al.* [22] using all available germline variants (including non-PASS) of the 1000 Genomes Project (phases 1 and 3) [23, 24], the non-TCGA Exome Aggregation Consortium (ExAC) database (r0.3.1) [25], and 11 large dbGap studies [26] that are not related to cancer (phs000254, phs000291, phs000473, phs000474, phs000653, phs000655, phs000687, phs000693, phs000711,

phs000744, and phs000827). Without normal matched samples, we were not able to distinguish somatic from private or rare germline variants, so they were analyzed together to identify potential driver mutations. Indels were excluded from analysis, as indel calls are more difficult to make, and the use of a single caller may introduce false-positive data.

E. Computational Analysis of the Genetic Variants

The functional impact of each variant was estimated by the Evolutionary Action (EA) method. The EA scores of human variants are available from the EA server at http://mammoth.bcm.tmc.edu/uea/. The EA of a missense mutation is calculated as a product of two terms that represent the evolutionary sensitivity of the mutated position and the magnitude of the amino acid substitution, both of which can be estimated from evolutionary data [27]. The final EA score ranges from 0 (no fitness change) to 100 (complete loss of function), such that they indicate the predicted percent of fitness loss.

To assess the separation of somatic or private germline variants from known germline variants in PC, we used purifying selection pressure λ [28]. Germline variants have been under stronger selection pressure than somatic variants, resulting in substantially steeper exponentially decaying distributions of EA scores. The coefficient λ of the decays, calculated through regression analysis on 10 equidistant bins of EA values, can be used to assess the quality of germline and somatic mutations [28]. As a reference, we used somatic variants from 20 cancer types (downloaded in October 2014 as 54 mutation annotation format files from https://tcga-data.nci.nih.gov) and germline variants from the 1000 Genomes Project (phase 3 dated May 2015 downloaded from ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/).

For variants of specific genes or pathways, we identified selection patterns by comparing the distributions of EA scores of variants with random nucleotide changes on the corresponding sequence using the Kolmogorov-Smirnov (K-S) test. The K-S test is a nonparametric test of the equality of continuous, one-dimensional probability distributions. The P values were calculated for the null hypothesis that the sample was drawn from the reference distribution, quantifying the maximum distance between the cumulative distributions.

2. Results

A. Quantity and Quality of DNA Extraction

A total of 31 unique FFPE samples were subjected to successful DNA extraction with a median yield of 7.8 μg/sample [interquartile range (IQR) 4.2 to 16.0] and a median A260/280 ratio of 2.0 (IQR 1.9 to 2.01). FFPE samples ranged from 1 to 26 years of storage. Two of 31 tissues (7%) had poor DNA quantity and quality and were excluded from analysis. The remaining 29 unique patient samples comprise the study cohort and underwent successful sequencing defined as having a nonfailed final library >200 bp (median 295 bp; IQR 205 to 368) covering ≥50% of the target at 20 times. Of these 29 patients, 16 patients had primary sporadic PC that was adequately treated with initial surgical resection, 6 patients had recurrent local disease, and 7 patients had metastatic PC (4 patients with lung metastases, 1 patient with bone metastasis, 1 patient with liver metastasis, and 1 patient with both bone and lung metastases). Their clinical characteristics are briefly summarized in Table 1 [27, 29]. The median ratio of transitions to transversions, a measure of DNA chemical cross-linking and modification, was 2.09 (IQR 2.06 to 2.11), well within the expected range of high-quality FFPE-extracted DNA and comparable to ratios reported in fresh-frozen tissues [17]. Genomic data was deposited into the National Center for Biotechnology Information Database of Genotypes and Phenotypes (NCBI dbGaP) [30].

B. Validity of Somatic Calls

Without normal matched samples, we were unable to distinguish between somatic and rare germline variants, so the number of these variants might be higher than the number of the

Table 1. Clinical and Pathologic Diagnostic Features of Patients With Sporadic PC Who Comprise the Study Cohort With Somatic or Private Germline Variants Called, Allelic Fraction, and Evolutionary Action Score

Patient Identification Number	Clinical Course	WHO Diagnostic Criteria	Tumor Size, cm	Current Status	Serum Ca at Dx	PTH at Dx	Gene	Variant	Allele Fraction	EA
001	Primary PC	Vascular invasion	1.6	DOC	12.3	1000	TP53	K132N	0.371	92.22
002	Recurrent PC	Vascular invasion, soft tissue extension	8	DOC	Unknown	Unknown	KMT2B	R1771Q	0.259	99.97
							SUN2	W582C	0.833	90.88
							NUP107	S167L	0.706	83.85
							TP53	R181C	0.827	54.85
							XAB2	D253N	0.406	53.53
003	Primary PC	Soft tissue extension	2.5	NED	14.3	507	LATS2	K793M	0.333	86.19
							SYNE1	I3456M	0.563	40.29
004	Recurrent PC	Vascular invasion, soft tissue extension	2.5	DOD	16	1200	CDC73	W32X	0.817	STOP
005	Metastatic PC	Vascular invasion, soft tissue extension, lung metastasis	5	AWD	13.4	659	MSH2	Q545X	0.846	STOP
							GLI3	Q710X	0.286	STOP
							NF1	R1748X	0.756	STOP
							POLR2E	A102V	0.395	57.98
006	Primary PC	Vascular invasion, soft tissue extension	1.5	DOD	13.6	143	VCAN	G3102S	0.578	78.18
							ERC1	N266S	0.549	44.06
007	Recurrent PC	Soft tissue extension	2.1	NED	16	700				
008	Metastatic PC, lung	Vascular invasion, soft tissue extension, lung metastasis	5	AWD	16	1256	ARID1B	M923K	0.388	93.07
							RFC5	R215T	0.323	59.82
009	Primary PC	Vascular invasion, soft tissue extension	3.5	NED	12.1	120	KMT2C	R4523S	0.214	77.69
010	Primary PC	Soft tissue extension	2	NED	15.3	536				
011	Primary PC	Soft tissue extension	1.8	NED	11.3	800				
012	Primary PC	Vascular invasion	1	NED	10.9	67	MEN1	G230X	0.88	STOP
							BRCA2	S3133L	0.481	79.58
							FANCL	L254V	0.436	53.8
013	Recurrent PC	Vascular invasion, soft tissue extension	3.5	AWD	17	1707	KMT2D	R2830Q	0.404	59.73
014	Metastatic PC	Vascular invasion, soft tissue extension, liver metastasis	4	AWD	13.2	298	CENPF	R3094X	0.324	STOP
							TSC1	R228X	0.638	STOP
							TP53	R306X	0.806	STOP
							PTPRB	R1844W	0.292	78.17
							POLR2L	A34T	0.22	66.12
							RAD50	S1244C	0.264	60.21
							RAD50	I1227M	0.222	55.23
015	Metastatic PC	Soft tissue extension, lung metastasis, bone metastasis	3	DOD	18	620	SYCP2	P495L	0.318	87.9
016	Primary PC	Soft tissue extension	2.3	NED	10.6	256	BRAF	G469A	0.4	61.77
017	Primary PC	Vascular invasion, soft tissue extension	3	NED	11.5	359				
018	Metastatic PC	Vascular invasion, soft tissue extension, bone metastasis	2.5	DOD	14.5	2801	ATM	L1327X	0.423	STOP
							AKAP9	E341V	0.511	46.94
019	Primary PC	Vascular invasion, soft tissue extension	4	NED	12	140	MEN1	H438P	0.822	93.03
020	Primary PC	Vascular invasion, soft tissue extension	4	NED	20.5	2203	PSMC3IP	K128X	0.539	STOP
							SYNE1	E5956G	0.528	77.53
021	Primary PC	Soft tissue extension	3.2	NED	11.8	114				
022	Primary PC	Soft tissue extension	2	NED	12.6	137				
023	Metastatic PC	Vascular invasion, soft tissue extension, lung metastasis	1.5	AWD	13.7	698	CDC73	R76X	0.413	STOP
							CENPF	S1780X	0.237	STOP
							CTNNB1	Q72X	0.278	STOP
							KDM5C	G536R	0.721	95.87
							CDC73	Y55C	0.377	89.54
024	Metastatic PC	Vascular invasion, soft tissue extension, lung metastasis	3	DOD	15.3	7200				
025	Primary PC	Soft tissue extension	Unknown	NED	13.5	213				

(*Continued*)

Table 1. Clinical and Pathologic Diagnostic Features of Patients With Sporadic PC Who Comprise the Study Cohort With Somatic or Private Germline Variants Called, Allelic Fraction, and Evolutionary Action Score (Continued)

Patient Identification Number	Clinical Course	WHO Diagnostic Criteria	Tumor Size, cm	Current Status	Serum Ca at Dx	PTH at Dx	Gene	Variant	Allele Fraction	EA
026	Recurrent PC	Soft tissue extension	3.7	AWD	Unknown	133	KIAA1549	W1853X	0.351	STOP
027	Primary PC	Vascular invasion, soft tissue extension	3.5	NED	12.1	Unknown	AKAP9	V1595L	0.473	54.46
028	Recurrent PC	Soft tissue extension	1.1	AWD	Unknown	Unknown	CDC73	Y55X	0.311	STOP
							ERBB4	Q1260X	0.447	STOP
029	Primary PC	Vascular invasion, soft tissue extension	2	DOC	13.2	567				

Clinical Course: primary PC was defined as parathyroid cancer by WHO diagnostic criteria without tumor recurrence after resection, recurrent PC was defined as local recurrence after curative resection, and metastatic PC was defined by pathologic or imaging confirmation of distant metastasis. WHO Diagnostic Criteria: the WHO states that the diagnosis of PC should limited to tumors with evidence of invasion into adjacent soft tissue, organs, or structures, invasion into capsular or extracapsular blood vessels or perineural spaces, and to patients with documented metastasis [29]. Tumor Size: primary parathyroid tumor size at diagnosis reported in centimeters. Current Status: alive with disease (AWD); died of other causes (DOC); died of disease (DOD); and alive with no evidence of disease (NED). Serum Ca at Dx: serum calcium (Ca) levels at diagnosis (Dx) reported in milligrams per deciliter. PTH at Dx: serum parathyroid hormone (PTH) level at Dx reported in picograms per milliliter. Allele Fraction: the relative frequency of the gene variant expressed as a fraction of the tumor chromosomes that that carry that allele. EA: a measure of fitness effect of the variants in each gene as calculated by the EA equation [27].

true somatic variants. As such, the mutations are reported as PC somatic or private germline mutations. We assessed for each sample from our cohort whether the number of somatic or private germline mutations and their distribution of fitness effects were consistent with observations from other cancer types and from two previously published cohorts of PC samples, the PC dataset from Yu *et al.* [14] that was obtained from Pandya *et al.* [16]. In our 29 samples, we found a range of 18 to 144 missense mutations per sample. This number of missense variants is comparable to that seen in other datasets, and the lambdas are similar to those of TCGA, suggesting that our data are consistent with what is expected and that the analysis is valid (Fig. 1A). The sample median was 49 missense mutations, which was intermediate to the medians of the two smaller published cohorts that had 21 and 84 missense mutations, respectively. Also, both extreme samples of our cohort were within the ranges of

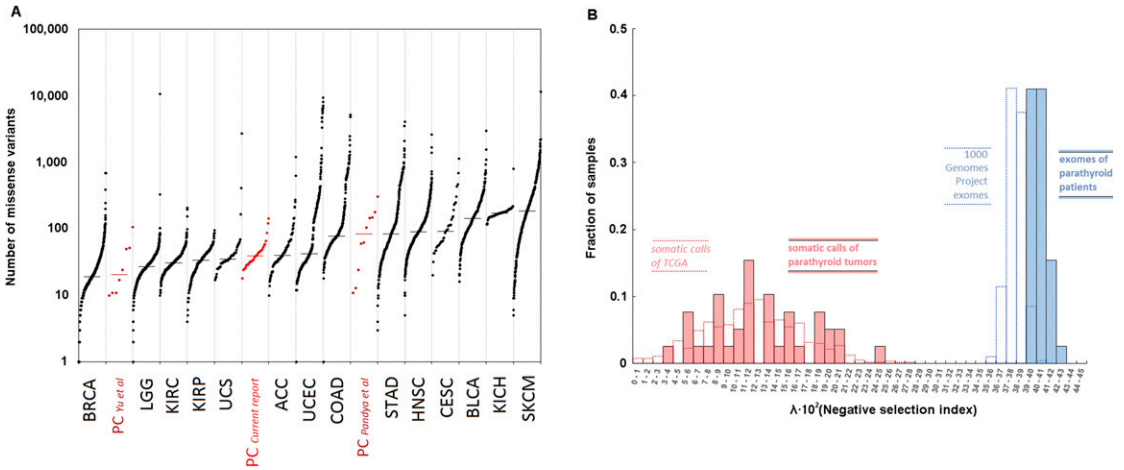

Figure 1. Somatic variant calls. (A) The average number of somatic missense mutations per sample for the reported PC cohort (red), the previously published PC cohorts (brown), and 14 cohorts of other cancer types (black) obtained from the TCGA portal. (B) The selection constraints index λ for somatic (red) and germline (blue) calls of the reported PC cohort (solid bars), compared with somatic variants of TCGA and germline variants of the 1000 Genomes Project, respectively (dotted bars).

the other two parathyroid cohorts. To estimate the extent of germline variants in our somatic calls, we calculated the purifying selection pressure (λ) [28]. The missense somatic or private germline mutation calls in our cohort had a distribution of λ values that was indistinguishable from TCGA variants (Fig. 1B) and from the somatic calls in the published PC cohorts. In contrast, the distributions of λ values for the variants we called germline were within and higher than the range for germline variants from the 1000 Genomes Project, which is higher and not overlapping with the range of somatic variants. These data show that even without matched normal samples, we could identify a group of somatic and private germline PC variants that have consistent features with the somatic variants of two prior PC cohorts and other cancer types. These variant calls and their corresponding allelic fractions (AFs) are summarized per patient in Table 1 [27, 29].

C. Candidate PC Driver Genes Identified

We tested whether PC somatic and private variants can indicate genes under positive selection with no bias to the current knowledge of cancer driver genes. We estimated the fitness effect of the variants in each gene with the EA equation [27] and asked whether the distribution of EA scores is different from the distribution of random nucleotide changes. We considered 85 genes that harbored at least two somatic variations in our PC cohort, ignoring 1178 genes that had only a single nonsynonymous variation. Seven genes had a P value ≤ 0.05 by K-S test; only *CDC73* remained significant after accounting for multiple testing (Supplemental Table S1). *CDC73* and *MEN1* are the only genes associated with parathyroid tumors based on somatic mutation data, according to TCGC [6], which were prioritized as first and fifth in this analysis (K-S P = 0.002 and 0.03, respectively). *C9orf37*, *AGFG2*, *CENPF*, and *FSIP2* had P values <0.05. Of these, centromere protein F (*CENPF* gene) plays a critical role in chromosome segregation during mitosis [31], and its expression levels have been associated with prostate cancer [32], hepatocellular carcinoma [31], and esophageal squamous cell carcinoma [33], suggesting it could be also implicated in PC. The genes with the most mutations in the parathyroid tumors (RGS3 and OSTC with six mutations each) had no bias to high impact ($P > 0.8$), and they have not been associated with any cancer type in TCGC. These data suggest that our approach to consider the fitness impact of somatic mutations can help in prioritizing candidate genes, although the discovery of PC driver genes might be limited by the small number of somatic mutations per gene.

Next, we asked whether genes with nonsignificant K-S P values may yet have potential to be driver genes of PC based on nonrandom functional relationships with top prioritized genes. A total of 39 genes with a P value <0.5 were significantly enriched in mutual protein-protein interactions according to the String v10.5 database ($P = 1.9 \times 10^{-3}$ for links with confidence ≥ 0.15; Fig. 2A), indicating a significant functional relationship among these genes. When 20 additional genes with a bias to high or intermediate EA scores typical of tumor suppressor and oncogene mutations were included in network interaction analysis, the enrichment in gene interactions remained significant ($P = 4 \times 10^{-4}$ for links with confidence ≥ 0.15; Fig. 2B). In addition to *CDC73* and *MEN1*, several genes strongly interacted (Fig. 2), including seven genes found in TCGC [(*TP53*, *AKAP9*, *KIAA1549*, *TSC1*, *ERC1*, *KMT2D* (*MLL2*), and *KMT2C* (*MLL3*)] and six others (*RAD50*, *CENPF*, *NUP107*, *SYNE1*, *LATS2*, and VCAN), suggesting a potential role of these genes in parathyroid carcinogenesis.

D. Pathway Analysis and Known Driver Selection

Using pathways from the Reactome database v49 [34, 35], we grouped the mutations of 7498 genes into 1580 pathways and their possible subpathways. We compared EA score distributions of each mutation set to random nucleotide changes and calculated the P values and the corrected for multiple test q values for each mutation set. The three highest-ranked sets of mutations implicated a total of 24 mutations across 16 genes that were key regulators of the β-catenin–TCF transactivating complex, meiosis, and DNA repair pathways (Fig. 3). These

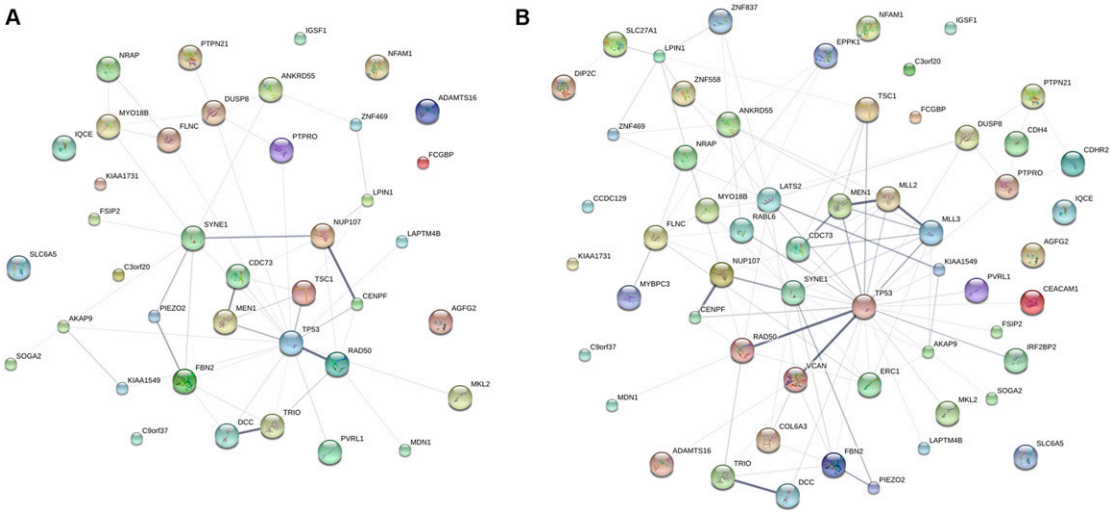

Figure 2. Protein interactions according to the String v10.5 database. The strength of the links indicates the confidence of association. Links with confidence ≥0.15 are shown. (A) Thirty-eight of the 39 genes with K-S *P* value ≤0.5 were able to map on the String network. (B) Fifty-seven of 59 genes with bias to high or intermediate EA scores were able to map on the String network.

16 genes were enriched for TCGC members ($P = 1.2 \times 10^{-5}$) and positively identified the known PC drivers *CDC73* and *MEN1* in the top-ranked pathway. The DNA repair gene set, comprised of nine genes including *ATM*, *BRCA2*, *MSH2*, and *RAD50*, genes known to be drivers in other cancer types [36–39] but not previously associated with PC. Twelve of the 16 genes were affected by only a single mutation in the cohort, three genes (*MEN1*, *SYNE1*, and *RAD50*) were affected by only two mutations, and one gene (*CDC73*) by four mutations; however, when viewed as a whole, these three pathways were mutated in 45% of the patients. These data suggest that these functionally related genes likely represent PC drivers.

We further tested whether grouping the mutations of known cancer driver genes can reveal selection patterns. We considered 213 genes that have been associated with any cancer type because of their somatic mutations, according to TCGC [6] (Supplemental Table S2). In our PC cohort, we observed 10 silent (S), 30 missense (M), and 11 nonsense (N) mutations in these 213 genes, compared with 559 S, 1289 M, and 54 N mutations in the rest of the genes (Supplemental Table S3). The increase in N/S ratio was >10-fold (Fig. 4A), suggesting that almost all CGC genes with any nonsense mutations, *CDC73*, *MEN1*, *TP53*, *NF1*, *ATM*, *TSC1*, *CTNNB1*, *ERBB4*, and *MSH2*, are likely PC driver candidates.

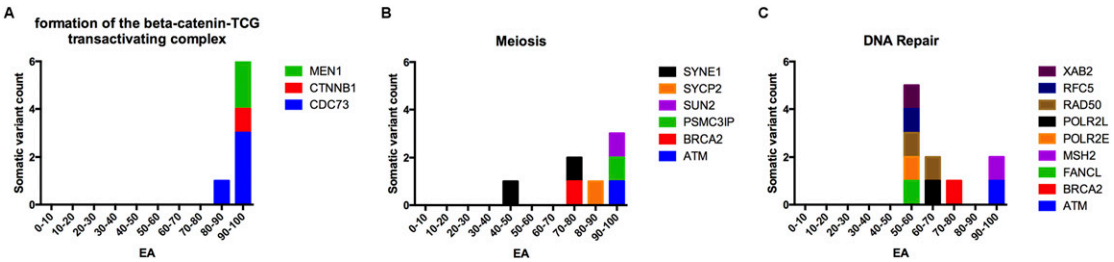

Figure 3. Gene pathways under positive selection in PC. (A) Formation of the β-catenin, (B) meiosis, and (C) DNA repair gene pathways of the Reactome database were found to be nonrandomly mutated in patients with PC. Nonsynonymous mutations are binned in deciles by their EA score. Higher scores indicate loss of function, whereas intermediate scores indicate gain of function or function separation. Absolute counts of mutations in each bin are represented by the height of the bin. Each component gene is represented by a color, and plots are labeled with the Reactome pathway associated with the gene set.

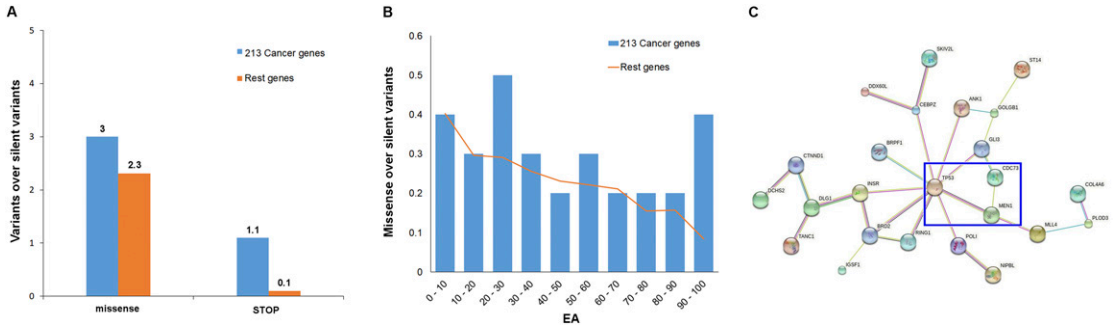

Figure 4. Somatic mutations of known cancer drivers show selection patterns in parathyroid tumor variants. (A) The ratios of missense-to-silent and nonsense-to-silent somatic parathyroid tumor variants for 213 cancer-associated genes (blue) and for the rest of the genes (orange). (B) The distribution of EA fitness scores for missense variants normalized by the number of silent mutations for the 213 cancer-associated genes (blue bars) and for the rest of the genes (orange line). (C) Gene-interaction links (String v10.5 database; link confidence >0.4) among *TP53*, *CDC73*, and *MEN1* with genes that have impactful variants in parathyroid tumors but they were not listed in TCGC.

We also compared the AF of the potential candidate drivers in Table 1 [27, 29], the passenger somatic or private variants, and the germline variants to further support our proposed candidate driver variants (Fig. S1). The N/S ratio of CGC genes was increased by 43% compared with the rest of the genes, and this bias was mostly due to missense mutations with high EA (Fig. 4B), which occurred on genes *TP53*, *MEN1*, *ARID1B*, *KDM5C*, *PTPRB*, and *CDC73*. Common oncogenic mutations have been listed as hotspot driver mutations in COSMIC, and we only found one such mutation: BRAF G469A (EA score of 62). To identify potential candidate driver genes among those not listed in TCGC, we considered genes with nonsense or high EA mutations in parathyroid tumors, and they are associated to *CDC73*, *MEN1*, and *TP53*. From 21 genes that meet these criteria (Fig. 4C), *GLI3* (Q710X) and *MLL4* (R1771Q) are directly linked to *CDC73* and *MEN1*, respectively, so they too may be specific candidate drivers of PC.

E. Germline Variant Analysis

To identify germline variants of patients with PC, we considered their relative impact compared with the general population. We used ExAC data [25] as reference. Gene impact was measured as a percentile rank of the mutation burden defined as the average EA score of the gene variants divided by the logarithm of the variant allelic frequency. Genes with the largest difference in mutation burden for the two cohorts (≥60%) were identified as potential PC-associated genes through germline variants. Despite the small number of PC germline variants (three to four variants per gene on average), genes impaired more severely in PC than in the general population were enriched in the Kyoto Encyclopedia of Genes and Genomes pathway annotations of MAPK signaling, T-cell receptor signaling, chronic and acute myeloid leukemia, and pathways in cancer (*P* values 0.002 to 0.02; Fig. 5A). In contrast, genes impaired less severely in PC than in the general population were enriched in immune response pathways, including a Kyoto Encyclopedia of Genes and Genomes annotation of autoimmune thyroid disease (*P* value of 10^{-7}; Fig. 5B). Together, these data suggest that germline variants may play a role in developing PC.

3. Discussion

Our understanding of PC carcinogenesis and the role of genetic and epigenetic factors in its tumor formation and progression have been severely limited by the low incidence of disease, with <100 new cases per year diagnosed in the United States [1, 40]. Although the TCGA and

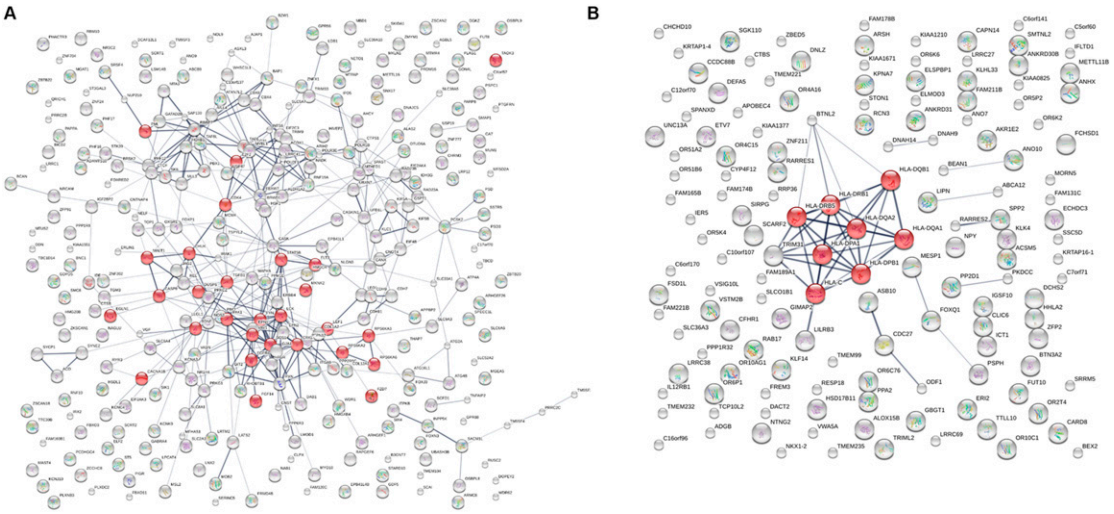

Figure 5. Genes with different germline mutation burden between patients with parathyroid cancer and general human population. (A) Genes that harbor germline variants with higher impact in patients with PC than in individuals of the ExAC database. (B) Genes that harbor germline variants with lower impact in patients with parathyroid cancer than in individuals of the ExAC database.

others have amassed large volumes of genetic data on a variety of tumors, PC has been largely ignored in efforts to study the genomic basis of cancer due to poor availability of fresh-frozen tissue. In this study, we present a genomic study of PC with whole-exome sequencing of 29 sporadic PC tumors using DNA extracted from FFPE tumor specimen with 93% successful whole-exome sequencing assay results and with acceptable DNA quantity and quality. Furthermore, the stated transition to transversion ratio suggests that fixation artifacts resulting in C>T substitutions are unlikely to have a major impact on our genomic findings. We identified potential candidate driver somatic variants and genes, and we compared the results with two independently published cohorts of PC totaling 18 exomes [14, 16]. We prioritized somatic and private germline variants with the EA equation and compiled a list of 35 genes that were prominent in one or more of the analyses we performed (single-gene analysis and gene interactions, gene pathway analysis, impactful mutations of TCGC genes, and interactors of *CDC73* and *MEN1*) (Table 2). Intriguingly, this table lists a considerable number of large genes that are often present in recurrent analyses of other cancer types. Although false positives are expected to be seen in large genes due to higher probability for mutation, the performed analyses, especially the nonsense mutations and pathways analyses, also favor the discovery of these large genes. Only 8 of these 35 genes were previously reported with any somatic mutation in previous PC cohorts, of which only 5 genes (*CDC73*, *AKAP9*, *NUP107*, *SYNE1*, and *VCAN*) had large enough EA impact (nonsense, frameshift indels, and missense with EA >30) as expected for driver mutations.

CDC73 variants were underrepresented in the current study, with only three patients found to have truncating CDC73 mutations. This may be because indel data were excluded from our analysis. The analysis was limited to single nucleotide variant data because indel calls are more difficult to make, and the use of a single caller may introduce false-positive data. Additionally, exon 1 of the CDC73 gene is known to have large GC repeats that are not well covered with NGS. Exon 1 is known to harbor a notable proportion of known CDC73 somatic mutations [13]. Other studies have identified five frameshift indels of CDC73 of 18 cases, suggesting that a substantial number of CDC73 mutations were not captured in our study and represent a considerable limitation. Further Sanger analysis of exon 1 may result in more calls. Notably, no *TP53* mutations were reported in either prior reported PC cohorts. Additionally, no somatic or private germline mutations were found in genes PRUNE2, MTOR, PIK3CA, and CCND1 in our study cohort. Although copy-number variations in CCND1 have been previously reported in PC, we did not evaluate copy number in the current

Table 2. Potential Candidate Driver Genes of Sporadic Parathyroid Cancer, Their Functional Pathways, and the Somatic or Private Germline Mutations Observed in the Current Cohort and Somatic Variants From Previous Studies

Gene	Function Pathways (PathCards)	Somatic Variants, Study Cohort of N = 29 (EA)
AKAP9	RET signaling and Regulation of PLK1 Activity	**E341V (**47**)**; **V1595L (54)**; T2427T
ARID1B	Activation of the ESR1/SP pathway	**M910K (88)**
ATM	DNA Double-Strand Break Repair	**L1327X**
BRAF	Trk receptor signaling and mTOR Pathway	**G496A (62)**
BRCA2	DNA Double-Strand Break Repair	**S3133L (80)**
CDC73	Signaling by Wnt and Signaling by Hedgehog	**W32X**; **Y55X**; **R76X**; **Y55C (90)**
CENPF	FOXM1 transcription factor and Regulation of PLK1 Activity	**S1780X**; **R3094X**
CTNNB1	Development VEGF and Endothelin-1/ EDNRA signaling	**Q72X**
ERBB4	GPCR Pathway and RET signaling	**Q1260X**
ERC1	IL-1 and NF-κB signaling pathways	**N266S (**44**)**; S969Y (29)
FANCL	Fanconi anemia and DNA repair	**L254V (54)**; F36F
GLI3	Wnt Signaling and Hedgehog Pathway	**Q710X**
KDM5C	Activated PKN1 and Chromatin organization	**G536R (96)**
KIAA1549	HIV Life Cycle and Oncogenic MAPK signaling	**W1853X**; A905T (27)
KMT2B (MLL4)	Lysine degradation and PKMTs methylate histone lysines	**R1771Q (100)**
KMT2C (MLL3)	Lysine degradation and Activated PKN1	L3483S (24); **R4523S (78)**
KMT2D (MLL2)	Lysine degradation and Signaling by Wnt	P610A (19); **R2830Q (60)**
LATS2	Hippo signaling and Signaling by GPCR	A428T (22); **K793M (86)**
MEN1	Transcriptional activity of SMADs and Signaling by Wnt	**G230X; H438P (93)**
MSH2	Mismatch repair and Platinum drug resistance	**Q545X**
NF1	Development VEGF signaling and RET signaling	**R1748X**
NUP107	HIV Life Cycle and Transport of the SLBP mRNA	**S167L (84)**; **L301M (**39**)**
POLR2E	HIV Life Cycle and Formation of HIV elongation complex	**A102V (58)**
POLR2L	HIV Life Cycle and Formation of HIV elongation complex	**A34T (66)**
PSMC3IP	Meiosis and Cell Cycle, Mitotic	**K116X**
PTPRB	Innate Immune System and Activation of cAMP-Dependent PKA	**R1754W (91)**
RAD50	DNA Double-Strand Break Repair	**I1227M (55); S1244C (60)**; L1211L; I1212I
RFC5	DNA Double-Strand Break Repair and Translesion synthesis	**R215T (60)**
SUN2	Cell Cycle, Mitotic and Meiosis	**W582C (91)**
SYCP2	Cell Cycle, Mitotic and Meiosis	**P495L (88)**
SYNE1	Cell Cycle, Mitotic and Meiosis	**I3456M (**40**)**; **E5956G (78)**; L3447L
TP53	IGF-1 receptor signaling and DNA Double-Strand Break Repair	**R306X; K132N (92)**; **R181C (55)**; T125T
TSC1	IGF-1 receptor signaling and RET signaling	**R177X;** V25M (10)
VCAN	Phospholipase-C Pathway and Chondroitin sulfate metabolism	P2996T (32); **G3102S (78)**
XAB2	mRNA Splicing and TC Nucleotide Excision Repair	**D253N (54)**

The mutual interactions of the genes are according to STRING v10.5 (for link confidence >0.15). Bold text indicates variants predicted to have high impact (EA score is above 30).
Abbreviations: GPCR, G-protein–coupled receptor; PKA, protein kinase A; PKMT, protein lysine methyltransferase; TC, transcription coupled; VEGF, vascular endothelial growth factor.

study, and so the impact of CCND1 in this PC cohort remains unknown [14–16]. *TSC1* and *NF1* variants were observed, each with AFs >0.5 (0.638 and 0.756, respectively), suggesting loss of heterozygosity. Both *TSC1* and *NF1* are critical negative regulators of *mTOR* with loss of heterozygosity leading to *mTOR* signaling activation and proliferative dysregulation [41, 42]. The presence of these two somatic variants further implicate the PIK3CA-mTOR-AKT axis in sporadic PC tumorigenesis and support the study and use of available therapeutics that target these pathways (such as everolimus) in this patient population.

CTNNB1 was also found to be mutated in our study. *CTNNB1* is known to be an oncogene, and it is therefore surprising to have a truncation mutation seen in PC. The *CTNNB1* variant Q72X was called in both sequenced samples of that individual with 15 out of 54 and 23 out of 57 reads supporting the variant call. *CTNNB1* truncations have been observed in 52 of 7307 tumor samples according to the COSMIC database. Although *CTNNB1* is a 781-amino-acid–long protein, these truncations were not randomly distributed in the sequence with the majority of the truncation mutations, 34 (65%) occurring within the first 100 amino acids (13% of the protein), including Q72X. This suggests that cancer cells may prefer to have an early sequence truncation of *CTNNB1* rather than a late one, for reasons we do not understand.

The differences in somatic variants observed between prior PC genomic cohorts illustrate the need for obtaining more sequencing data of parathyroid tumors to understand the underlying biological mechanisms that cause PC beyond the role of *CDC73* deactivation. Our list of 35 potential candidate genes is a major step toward this goal; however, further validation is needed. Some of these 35 potential PC candidate genes also contained impactful germline variant that might be disease risk factors, but, due to sample size and data structure limitations, did not yield notable associations. Notably, no previously reported germline variants were found in *CDC73*, and only three common silent variants were found in *MEN1*. The group of 35 genes had 181 nonsynonymous germline variants collectively, of which only 58 had sufficiently large EA impact (EA ≥30) to be considered as potential PC driver variants (Table 2). Perhaps of most interest are two *TP53* variants (R337C and R248W with EA scores of 63 and 89, respectively), both of which are frequent cancer driver mutations according to COSMIC [43] (although annotated as germline variants because of their allele frequency of 8×10^{-6} in ExAC, they may be somatic variants in patients with PC), and three *RAD50* variants (T191I, V315L, and R327H with EAs of 58, 61, and 42, respectively). Other genes with potentially interesting germline mutations include *ATM*, *GLI3*, *KIAA1549*, *KMT2B*, *KMT2C*, and *SUN2*. Beyond these 35 genes, we also considered those with higher or lower impact in PC compared with the general population as represented by the ExAC database.

We acknowledge several limitations in our study. First, matched germline control samples were not available for analysis given the retrospective nature of this study. For this reason, we could not distinguish between somatic and private germline mutations, and so these variants were analyzed together. Fresh-frozen tissue is the gold standard for DNA isolation and sequencing, but in this rare tumor, prospective studies to amass substantial numbers of fresh-frozen samples are not feasible without wide-scale national and international collaborations. Retrospective studies using available FFPE tissue offer a good alternative for genomic insight in this disease. The use of FFPE samples for high-throughput parallel sequencing has not been widely adopted due to concerns for high rates of nonreproducible DNA sequence alterations from formalin cross-linking causing artificial C>T transitions [44, 45]. However, several recent publications have reported feasibility and clinical utility of NGS on FFPE tissue using the methodology used in our present study with high concordance between matched fresh-frozen and FFPE with successful sequencing tied more to DNA extraction quantity and quality [17, 18, 46]. Carrick *et al.* [17] reported the ratio of transitions to transversions increased with storage time, these ratios were only marginally higher in FFPE samples, even in tissue stored up to 32 years, when compared with fresh-frozen tissue. Similarly, Spencer *et al.* [46] found high concordance between NGS genomic analysis from matched fresh-frozen and FFPE tissue and concluded that though there is a detectable effect on DNA after FFPE preservation, this effect has a negligible impact on NGS with appropriate extraction and sequencing optimization. In the current study, the median time of FFPE

tissue storage was 11 years (IQR 5 to 16), and DNA extraction resulted in acceptable transition to transversion ratios with low library failure rates (7%) and good coverage statistics.

We used a single-algorithm VarScan v2.3.9 to make variant calls. VarScan v2.3.9 is a well-validated pipeline variant calling that employs strategies to minimize false positives associated with common sequencing-related and alignment-related artifacts with high sensitivity and specificity [20]. However, no consensus calls were made using multiple pipelines, and manual review of each variant call was not performed. As such, there is potential for false-positive or false-negative calls in this analysis. Sanger analysis was not performed in this study, so that areas of the genome susceptible to be missed by NGS (*i.e.*, large deletions and high GC content) may not be well captured and represent another limitation of this study.

Though a sample size of 29 tumors is substantial for sporadic PC given its very low incidence, genes with low frequency that may carry significance in sporadic PC tumor initiation and progression are likely not captured in this analysis. Because sequencing was carried out in tumor samples only, and though common variants were filtered, it is possible some of the mutations we identified in the expansion cohort (Supplemental Table S1) are rare or private germline variants rather than somatic mutations. Due to restricted sample availability, we were unable to generate genomic data in other platforms such as gene expression profiling. The genetic alterations identified through exome sequencing would be even better interpreted if integrative data analysis could have been performed with genomic data based on multiple platforms, as demonstrated in TCGA. Gene expression data would allow us to corroborate and extend copy number variation analysis results such as CCND1 amplification. Nevertheless, the current study represents a robust genomic sequencing effort and analysis for an extremely rare malignancy. Our results have indeed revealed important insights into the genetic landscape of PCs. The phosphatidylinositol 3-kinase/AKT/mTOR pathway is again identified as a potentially major oncogenic pathway in sporadic PC. In our study, both *TSC1* and *NF1* variants have AFs >0.5 (0.638 and 0.756, respectively), which indicate loss of heterozygosity. This further implicates the PIK3CA/AKT/mTOR axis in sporadic PC tumorigenesis and adds to the growing body of data that support the use of everolimus in metastatic PC.

In summary, we studied the genetic variations of 29 PCs to identify candidate driver genes. Although currently this is the largest available cohort, the relatively small number of samples compared with other cancer types is a limiting factor to discover driver genes with high confidence. To enhance our ability to discover genes, we accounted for the fitness impact of the genetic variations using the EA equation. EA predictions were shown to correlate with experimental and clinical data better than other approaches [47] and have been used before to identify driver genes in hepatocellular carcinoma [48] and to separate patient survival in head and neck cancer [49] and colorectal liver metastases [50]. Our study identified, besides the known PC driver genes *CDC73* and *MEN1*, 33 additional candidates, most associated with other cancer types. Germline variants may predispose to PC, especially variants in genes associated with MAPK signaling, T-cell receptor signaling, and immune response. Our findings are tempered by the inherent molecular artifacts introduced during preservation of FFPE tissue and the lack of matched germline DNA available for analysis. Additionally, variant calls were made using only a single algorithm and as such, may include false-positive or false-negative calls. Nevertheless, these data provide insight into the biology of sporadic PC and open ground for further genetic and therapeutic studies to better treat this disease.

Acknowledgments

Financial Support: Research reported in this study was supported by the National Institutes of Health under Award Number T325T32CA009599-24, the Baylor College of Medicine Comprehensive Cancer Training Program Award Number RP160283, and an MDACC Parathyroid Cancer Research Grant. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Correspondence: Callisia N. Clarke, MD, Division of Surgical Oncology, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, Wisconsin 53226. E-mail: cnclarke@mcw.edu.

Disclosure Summary: The authors have nothing to disclose.

References and Notes

1. Hundahl SA, Fleming ID, Fremgen AM, Menck HR; The American College of Surgeons Commission on Cancer and the American Cancer Society. Two hundred eighty-six cases of parathyroid carcinoma treated in the U.S. between 1985-1995: a National Cancer Data Base Report. *Cancer*. 1999;**86**(3): 538–544.
2. Busaidy NL, Jimenez C, Habra MA, Schultz PN, El-Naggar AK, Clayman GL, Asper JA, Diaz EM Jr, Evans DB, Gagel RF, Garden A, Hoff AO, Lee JE, Morrison WH, Rosenthal DI, Sherman SI, Sturgis EM, Waguespack SG, Weber RS, Wirfel K, Vassilopoulou-Sellin R. Parathyroid carcinoma: a 22-year experience. *Head Neck*. 2004;**26**(8):716–726.
3. Christakis I, Silva AM, Kwatampora LJ, Warneke CL, Clarke CN, Williams MD, Grubbs EG, Lee JE, Busaidy NL, Perrier ND. Oncologic progress for the treatment of parathyroid carcinoma is needed. *J Surg Oncol*. 2016;**114**(6):708–713.
4. Silva-Figueroa AM, Hess KR, Williams MD, Clarke CN, Christakis I, Graham PH, Grubbs EG, Lee JE, Busaidy NL, Perrier ND. Prognostic scoring system to risk stratify parathyroid carcinoma [published online ahead of print 15 March 2017]. *J Am Coll Surg*. doi: 10.1016/j.jamcollsurg.2017.01.060.
5. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn)*. 2015;**19**(1A):A68–A77.
6. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR. A census of human cancer genes. *Nat Rev Cancer*. 2004;**4**(3):177–183.
7. Carpten JD, Robbins CM, Villablanca A, Forsberg L, Presciuttini S, Bailey-Wilson J, Simonds WF, Gillanders EM, Kennedy AM, Chen JD, Agarwal SK, Sood R, Jones MP, Moses TY, Haven C, Petillo D, Leotlela PD, Harding B, Cameron D, Pannett AA, Höög A, Heath H III, James-Newton LA, Robinson B, Zarbo RJ, Cavaco BM, Wassif W, Perrier ND, Rosen IB, Kristoffersson U, Turnpenny PD, Farnebo LO, Besser GM, Jackson CE, Morreau H, Trent JM, Thakker RV, Marx SJ, Teh BT, Larsson C, Hobbs MR. HRPT2, encoding parafibromin, is mutated in hyperparathyroidism-jaw tumor syndrome. *Nat Genet*. 2002;**32**(4):676–680.
8. Cetani F, Pardi E, Borsari S, Viacava P, Dipollina G, Cianferotti L, Ambrogini E, Gazzerro E, Colussi G, Berti P, Miccoli P, Pinchera A, Marcocci C. Genetic analyses of the HRPT2 gene in primary hyperparathyroidism: germline and somatic mutations in familial and sporadic parathyroid tumors. *J Clin Endocrinol Metab*. 2004;**89**(11):5583–5591.
9. Rawat N, Khetan N, Williams DW, Baxter JN. Parathyroid carcinoma. *Br J Surg*. 2005;**92**(11): 1345–1353.
10. Cinque L, Sparaneo A, Cetani F, Coco M, Clemente C, Chetta M, Balsamo T, Battista C, Sanpaolo E, Pardi E, D'Agruma L, Marcocci C, Maiello E, Hendy GN, Cole DEC, Scillitani A, Guarnieri V. Novel association of *MEN1* gene mutations with parathyroid carcinoma. *Oncol Lett*. 2017;**14**(1):23–30.
11. Howell VM, Haven CJ, Kahnoski K, Khoo SK, Petillo D, Chen J, Fleuren GJ, Robinson BG, Delbridge LW, Philips J, Nelson AE, Krause U, Hammje K, Dralle H, Hoang-Vu C, Gimm O, Marsh DJ, Morreau H, Teh BT. HRPT2 mutations are associated with malignancy in sporadic parathyroid tumours. *J Med Genet*. 2003;**40**(9):657–663.
12. Haven CJ, van Puijenbroek M, Tan MH, Teh BT, Fleuren GJ, van Wezel T, Morreau H. Identification of MEN1 and HRPT2 somatic mutations in paraffin-embedded (sporadic) parathyroid carcinomas. *Clin Endocrinol (Oxf)*. 2007;**67**(3):370–376.
13. Shattuck TM, Välimäki S, Obara T, Gaz RD, Clark OH, Shoback D, Wierman ME, Tojo K, Robbins CM, Carpten JD, Farnebo LO, Larsson C, Arnold A. Somatic and germ-line mutations of the HRPT2 gene in sporadic parathyroid carcinoma. *N Engl J Med*. 2003;**349**(18):1722–1729.
14. Yu W, McPherson JR, Stevenson M, van Eijk R, Heng HL, Newey P, Gan A, Ruano D, Huang D, Poon SL, Ong CK, van Wezel T, Cavaco B, Rozen SG, Tan P, Teh BT, Thakker RV, Morreau H. Whole-exome sequencing studies of parathyroid carcinomas reveal novel PRUNE2 mutations, distinctive mutational spectra related to APOBEC-catalyzed DNA mutagenesis and mutational enrichment in kinases associated with cell migration and invasion. *J Clin Endocrinol Metab*. 2015;**100**(2):E360–E364.
15. Kasaian K, Wiseman SM, Thiessen N, Mungall KL, Corbett RD, Qian JQ, Nip KM, He A, Tse K, Chuah E, Varhol RJ, Pandoh P, McDonald H, Zeng T, Tam A, Schein J, Birol I, Mungall AJ, Moore RA, Zhao Y, Hirst M, Marra MA, Walker BA, Jones SJ. Complete genomic landscape of a recurring sporadic parathyroid carcinoma. *J Pathol*. 2013;**230**(3):249–260.

16. Pandya C, Uzilov AV, Bellizzi J, Lau CY, Moe AS, Strahl M, Hamou W, Newman LC, Fink MY, Antipin Y, Yu W, Stevenson M, Cavaco BM, Teh BT, Thakker RV, Morreau H, Schadt EE, Sebra R, Li SD, Arnold A, Chen R. Genomic profiling reveals mutational landscape in parathyroid carcinomas. *JCI Insight*. 2017;**2**(6):e92061.
17. Carrick DM, Mehaffey MG, Sachs MC, Altekruse S, Camalier C, Chuaqui R, Cozen W, Das B, Hernandez BY, Lih CJ, Lynch CF, Makhlouf H, McGregor P, McShane LM, Phillips Rohan J, Walsh WD, Williams PM, Gillanders EM, Mechanic LE, Schully SD. Robustness of next generation sequencing on older formalin-fixed paraffin-embedded tissue. *PLoS One*. 2015;**10**(7):e0127353.
18. Hedegaard J, Thorsen K, Lund MK, Hein AM, Hamilton-Dutoit SJ, Vang S, Nordentoft I, Birkenkamp-Demtröder K, Kruhøffer M, Hager H, Knudsen B, Andersen CL, Sørensen KD, Pedersen JS, Ørntoft TF, Dyrskjøt L. Next-generation sequencing of RNA and DNA isolated from paired fresh-frozen and formalin-fixed paraffin-embedded samples of human cancer and normal tissue. *PLoS One*. 2014;**9**(5):e98187.
19. Chen K, Meric-Bernstam F, Zhao H, Zhang Q, Ezzeddine N, Tang LY, Qi Y, Mao Y, Chen T, Chong Z, Zhou W, Zheng X, Johnson A, Aldape KD, Routbort MJ, Luthra R, Kopetz S, Davies MA, de Groot J, Moulder S, Vinod R, Farhangfar CJ, Shaw KM, Mendelsohn J, Mills GB, Eterovic AK. Clinical actionability enhanced through deep targeted sequencing of solid tumors. *Clin Chem*. 2015;**61**(3):544–553.
20. Koboldt DC, Larson DE, and Wilson RK. Using VarScan 2 for germline variant calling and somatic mutation detection. *Curr Protoc Bioinformatics*, 2013;**44**(1):15.4.1–15.4.17.
21. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;**38**(16):e164.
22. Hiltemann S, Jenster G, Trapman J, van der Spek P, Stubbs A. Discriminating somatic and germline mutations in tumor DNA samples without matching normals. *Genome Res*. 2015;**25**(9):1382–1390.
23. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA; 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;**491**(7422):56–65.
24. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR; 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*. 2015;**526**(7571):68–74.
25. 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, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won HH, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG; Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;**536**(7616):285–291.
26. Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, Hao L, Kiang A, Paschall J, Phan L, Popova N, Pretel S, Ziyabari L, Lee M, Shao Y, Wang ZY, Sirotkin K, Ward M, Kholodov M, Zbicz K, Beck J, Kimelman M, Shevelev S, Preuss D, Yaschenko E, Graeff A, Ostell J, Sherry ST. The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet*. 2007;**39**(10):1181–1186.
27. Katsonis P, Lichtarge O. A formal perturbation equation between genotype and phenotype determines the Evolutionary Action of protein-coding variations on fitness. *Genome Res*. 2014;**24**(12):2050–2058.
28. Koire A, Katsonis P, Lichtarge O. Repurposing germline exomes of the Cancer Genome Atlas demands a cautious approach and sample-specific variant filtering. *Pac Symp Biocomput*. 2016;**21**:207–218.
29. Lloyd RV, Osamura RY, Klöppel G, Rosai J, eds. *WHO Classification of Tumours of Endocrine Organs*. 4th ed. Lyon, France: IARC Press; 2017.
30. Clarke CN, Katsonis P, Hsu T-K, Koire AM, Silva-Figueroa A, Christakis I, Williams MD, Kutahyalioglu M, Kwatampora L, Xi Y, Lee JE, Koptez ES, Busaidy NL, Perrier ND, Lichtarge O. Comprehensive genomic characterization of parathyroid cancer identifies novel candidate driver mutations and core pathways. dbGaP 2019. Deposited 5 February 2019. www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001765.v1.p1.
31. Dai Y, Liu L, Zeng T, Zhu YH, Li J, Chen L, Li Y, Yuan YF, Ma S, Guan XY. Characterization of the oncogenic function of centromere protein F in hepatocellular carcinoma. *Biochem Biophys Res Commun*. 2013;**436**(4):711–718.

32. Zhuo Y-J, Xi M, Wan YP, Hua W, Liu YL, Wan S, Zhou YL, Luo HW, Wu SL, Zhong WD, Wu CL. Enhanced expression of centromere protein F predicts clinical progression and prognosis in patients with prostate cancer. *Int J Mol Med*. 2015;**35**(4):966–972.
33. Pimkhaokham A, Shimada Y, Fukuda Y, Kurihara N, Imoto I, Yang ZQ, Imamura M, Nakamura Y, Amagasa T, Inazawa J. Nonrandom chromosomal imbalances in esophageal squamous cell carcinoma cell lines: possible involvement of the ATF3 and CENPF genes in the 1q32 amplicon. *Jpn J Cancer Res*. 2000;**91**(11):1126–1133.
34. Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, Caudy M, Garapati P, Gillespie M, Kamdar MR, Jassal B, Jupe S, Matthews L, May B, Palatnik S, Rothfels K, Shamovsky V, Song H, Williams M, Birney E, Hermjakob H, Stein L, D'Eustachio P. The Reactome pathway knowledgebase. *Nucleic Acids Res*. 2014;**42**(Database issue):D472–D477.
35. Fabregat A, Sidiropoulos K, Garapati P, Gillespie M, Hausmann K, Haw R, Jassal B, Jupe S, Korninger F, McKay S, Matthews L, May B, Milacic M, Rothfels K, Shamovsky V, Webber M, Weiser J, Williams M, Wu G, Stein L, Hermjakob H, D'Eustachio P. The reactome pathway knowledgebase. *Nucleic Acids Res*. 2016;**44**(D1):D481–D487.
36. 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**(11):813–822.
37. Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, Collins N, Gregory S, Gumbs C, Micklem G. Identification of the breast cancer susceptibility gene BRCA2 [published correction appears in *Nature*. 1996;**379**(6567):749]. *Nature*. 1995;**378**(6559):789–792.
38. Fishel R, Lescoe MK, Rao MR, Copeland NG, Jenkins NA, Garber J, Kane M, Kolodner R. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell*. 1993; **75**(5):1027–1038.
39. Heikkinen K, Rapakko K, Karppinen SM, Erkko H, Knuutila S, Lundán T, Mannermaa A, Børresen-Dale AL, Borg A, Barkardottir RB, Petrini J, Winqvist R. RAD50 and NBS1 are breast cancer susceptibility genes associated with genomic instability. *Carcinogenesis*. 2006;**27**(8):1593–1599.
40. Lee PK, Jarosek SL, Virnig BA, Evasovich M, Tuttle TM. Trends in the incidence and treatment of parathyroid cancer in the United States. *Cancer*. 2007;**109**(9):1736–1741.
41. Huang J, Manning BD. The TSC1-TSC2 complex: a molecular switchboard controlling cell growth. *Biochem J*. 2008;**412**(2):179–190.
42. Johannessen CM, Reczek EE, James MF, Brems H, Legius E, Cichowski K. The NF1 tumor suppressor critically regulates TSC2 and mTOR [published correction appears in *Proc Natl Acad Sci USA*. 2005; **102**(44):16119]. *Proc Natl Acad Sci USA*. 2005;**102**(24):8573–8578.
43. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, Cole CG, Ward S, Dawson E, Ponting L, Stefancsik R, Harsha B, Kok CY, Jia M, Jubb H, Sondka Z, Thompson S, De T, Campbell PJ. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res*. 2017;**45**(D1):D777–D783.
44. Williams C, Pontén F, Moberg C, Söderkvist P, Uhlén M, Pontén J, Sitbon G, Lundeberg J. A high frequency of sequence alterations is due to formalin fixation of archival specimens. *Am J Pathol*. 1999; **155**(5):1467–1471.
45. Srinivasan M, Sedmak D, Jewell S. Effect of fixatives and tissue processing on the content and integrity of nucleic acids. *Am J Pathol*. 2002;**161**(6):1961–1971.
46. Spencer DH, Sehn JK, Abel HJ, Watson MA, Pfeifer JD, Duncavage EJ. Comparison of clinical targeted next-generation sequence data from formalin-fixed and fresh-frozen tissue specimens. *J Mol Diagn*. 2013;**15**(5):623–633.
47. Katsonis P, Lichtarge O. Objective assessment of the evolutionary action equation for the fitness effect of missense mutations across CAGI-blinded contests. *Hum Mutat*. 2017;**38**(9):1072–1084.
48. Cancer Genome Atlas Research Network. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. *Cell*. 2017;**169**(7):1327–1341.
49. Neskey DM, Osman AA, Ow TJ, Katsonis P, McDonald T, Hicks SC, Hsu TK, Pickering CR, Ward A, Patel A, Yordy JS, Skinner HD, Giri U, Sano D, Story MD, Beadle BM, El-Naggar AK, Kies MS, William WN, Caulin C, Frederick M, Kimmel M, Myers JN, Lichtarge O. Evolutionary action score of TP53 identifies high-risk mutations associated with decreased survival and increased distant metastases in head and neck cancer. *Cancer Res*. 2015;**75**(7):1527–1536.
50. Chun YS, Passot G, Yamashita S, Nusrat M, Katsonis P, Loree JM, Conrad C, Tzeng CD, Xiao L, Aloia TA, Eng C, Kopetz SE, Lichtarge O, Vauthey JN. Deleterious effect of RAS and evolutionary high-risk TP53 double mutation in colorectal liver metastases [published online ahead of print 1 Aug 2017]. *Ann Surg*. doi: 10.1097/SLA.0000000000002450.