Profile of basal cell carcinoma mutations and copy number alterations - focus on gene-associated noncoding variants

Paulina Maria Nawrocka1, Paulina Galka-Marciniak1, Martyna Olga Urbanek-Trzeciak1, Ilamathi M.1, Natalia Szostak2, Anna Philips2, Laura Susok3, Michael Sand3,4, Piotr Kozłowski1,*

1. Department of Molecular Genetics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland
2. Laboratory of Bioinformatics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland
3. Department of Dermatology, Venereology and Allergology, St. Josef Hospital, Ruhr-University Bochum, Bochum, Germany
4. Department of Plastic Surgery, St. Josef Hospital, Catholic Clinics of the Ruhr Peninsula, Essen, Germany

*Corresponding author, Tel: +48-61-8528503; Fax: +48-61-8520532; Email: kozlowp{at}ibch.poznan.pl
ABSTRACT

Basal cell carcinoma (BCC) of the skin is the most common cancer in humans, characterized by the highest mutation rate among cancers, and is mostly driven by mutations in genes involved in the hedgehog pathway. To date, almost all BCC genetic studies have focused exclusively on protein-coding sequences; therefore, the impact of noncoding variants on the BCC genome is unrecognized. In this study, with the use of whole-exome sequencing of 27 tumor/normal pairs of BCC samples, we performed an analysis of somatic mutations in both protein-coding sequences and gene-associated noncoding regions, including 5'UTRs, 3'UTRs, and exon-adjacent intron sequences. Separately, in each region, we performed hotspot identification, mutation enrichment analysis, and cancer driver identification with OncodriveFML. Additionally, we performed a whole-genome copy number alteration analysis with GISTIC2. Of the >80,000 identified mutations, ~50% were localized in noncoding regions. The results of the analysis generally corroborated the previous findings regarding genes mutated in coding sequences, including PTCH1, TP53, and MYCN, but more importantly showed that mutations were also clustered in specific noncoding regions, including hotspots. Some of the genes specifically mutated in noncoding regions were identified as highly potent cancer drivers, of which BAD had a mutation hotspot in the 3'UTR, DHODH had a mutation hotspot in the Kozak sequence in the 5'UTR, and CHCHD2 frequently showed mutations in the 5'UTR. All of these genes are functionally implicated in cancer-related processes (e.g., apoptosis, mitochondrial metabolism, and de novo pyrimidine synthesis) or the pathogenesis of UV radiation-induced cancers. We also found that the identified BAD and CHCHD2 mutations frequently occur in melanoma but not in other cancers via The Cancer Genome Atlas analysis. Finally, we identified frequent deletion of chr9q, encompassing PTCH1, and unreported frequent copy number gain of chr9p, encompassing the genes encoding the immune checkpoint ligands PD-L1 and PD-L2. In conclusion, this study is the first systematic analysis.
of coding and noncoding mutations in BCC and provides a strong basis for further analyses of the variants in BCC and cancer in general.

Author summary

The study is the first systematic analysis of both coding and noncoding mutations in basal cell carcinoma (BCC), the most common and the most highly mutated human cancer. Noncoding mutations accounted for ~50% (~40K mutations) of all mutations detected by the standard WES approach. Among the genes frequently mutated in noncoding regions are: BAD with a hotspot in the 3’UTR, DHODH with a hotspot in the Kozak sequence, and CHCHD2 with mutations in 5’UTR, all genes functionally related to cell metabolism and apoptosis. Analysis of copy number alterations showed frequent chr9q deletion, encompassing PTCH1 (the key BCC tumor suppressor), and frequent copy number gain of chr9p, encompassing the genes of the immune checkpoint proteins PD-L1 and PD-L2.

Keywords:
Basal cell carcinoma (BCC), cancer somatic mutations, noncoding mutations, immune checkpoint, copy number alterations, cancer drivers
INTRODUCTION

Basal cell carcinoma (BCC), a type of nonmelanoma skin cancer (NMSC), is the most common human cancer affecting predominantly elderly people of the Caucasian population [1–3]. The lifetime risk of BCC in the Caucasian population is ~30%, and it is higher in men and fair-skinned people. BCC usually occurs sporadically but can also develop as a result of Gorlin syndrome (also known as nevoid basal cell carcinoma syndrome (NBCCS)), an autosomal dominant hereditary condition with an incidence of approximately 1:30,000 [4] characterized by the frequent appearance of multiple BCC lesions that develop at a younger age together with skeletal abnormalities, odontogenic keratocysts, and an increased risk of medulloblastoma. Histologically, BCCs are classified into three major subtypes: nodular, which is the most common subtype; superficial; and infiltrative or sclerodermiform. Other subtypes as well as mixed types occur less frequently [5–7]. Predominantly, superficial and nodular BCCs are slow-growing, locally invasive, epidermal tumors with a metastasis rate of <0.1% [8,9], while infiltrative BCCs are characterized by more aggressive, tong-like, subclinical growth patterns mimicking icebergs, as they often grow below clinically healthy-looking skin [10,11]. Although BCC aggressiveness overall is low and metastatic potential is negligible, the commonness of BCC and the increasing incidence associated predominantly with aging populations has brought attention to its pathogenesis [2,3,12–17]. Exposure to ultraviolet (UV) radiation, which can lead to point mutations frequently represented by C>T and CC>TT transitions, is the main causative factor in the pathogenesis of BCC [18]. Additional risk factors include ionizing radiation, arsenic ingestion, and immune suppression [19,20].

BCC is characterized by the highest mutation rate observed among cancers, having over 65 mutations/Mbp [14,15]. The most frequent genetic alterations occurring in BCC are mutations disturbing the hedgehog (SHH/PTCH1/SMO) pathway, predominantly loss-of-function mutations in PTCH1 but also activating mutations in SMO; these genes encode two
transmembrane proteins, PTCH1 (also known as Patched1) and SMO (also known as Smoothened), respectively [14,15]. The pathway is activated by the SHH signaling protein (also known as Sonic hedgehog), which binds to the extracellular domain of PTCH1, disabling inhibition of SMO; this in turn activates GLI transcription factors. Germline mutations in \textit{PTCH1} predispose patients to Gorlin syndrome [21].

Previous studies, including whole-exome sequencing (WES) analyses, have also recognized other genes/pathways frequently mutated in BCC, including \textit{TP53}, \textit{MYCN}, \textit{PPP6C}, \textit{PTPN14}, \textit{STK19}, and \textit{LATS1} [14,15], as well as genes involved in the RTK-RAS-PI3K and Hippo-YAP pathways [15]. However, as an overwhelming majority of BCC genetic studies (as well as those in other cancers) have focused almost exclusively on protein-coding sequences, very little is known about mutations in noncoding regions (noncoding mutations). Noncoding mutations are not studied/reported even if detected, e.g., as a result of WES. On the other hand, it is well known that the noncoding parts of genes, i.e., promoters, introns, or 5' and 3' untranslated regions (5'UTRs and 3'UTRs, respectively), encompass numerous functional elements important for the proper functioning of the genes [22–24]. Somatic mutations may disrupt or modify the properties of these elements, acting either as gain- or loss-of-function mutations and thus enhancing/accelerating or switching off the function of some genes. Despite the negligence of the analysis of noncoding regions, there are some spectacular examples of noncoding driver mutations, for example, TERT promoter mutations, which occur most frequently in melanoma brain and bladder cancers but are also reported in BCC [25–27], and mutations in the precursor of miR-142, which frequently occur in Hodgkin lymphomas and acute myeloid leukemia (summarized in [28]). The miRNA biogenesis enzyme DICER has also been shown to bear mutations that could play a role in aberrant miRNA expression in BCC [29–31]. It should also be noted that an effort to catalog cancer somatic mutations in the
noncoding genome has recently been undertaken [32,33]; however, this pancancer project does not include BCC.

To preliminarily explore the occurrence of noncoding somatic mutations in BCC, we performed WES of over two dozen BCC samples, extending the analysis beyond protein-coding sequences and focusing on gene-associated noncoding regions, i.e., 5'UTRs, 3'UTRs, and exon-adjusted sequences of introns. Apart from the fact that our results well-replicate those of previous BCC studies in terms of mutations in protein-coding genes, we showed that a substantial portion of mutations are located in noncoding regions. Many of these mutations frequently recur in particular noncoding regions or in specific hotspot positions. Computational analyses showed that some of the gene mutations in noncoding regions are potential cancer drivers and are functionally related to skin cancers. Additionally, whole-genome copy number alteration (CNA) analysis revealed frequent deletion of chr9q, encompassing \textit{PTCH1}, and unreported frequent amplification of chr9p, including the genes encoding two immune checkpoint ligands PD-L1 and PD-L2.
RESULTS

Overall sequencing and mutation occurrence characterization

We performed WES on 27 paired tumor and corresponding intraindividual control skin DNA samples isolated from 22 nodular and 5 superficial BCC subtypes and corresponding healthy skin tissue. The average coverage of the targeted regions was 183x (185x in normal and 180x in tumor samples), ranging in different samples from 134x to 232x. In total, we identified 84,571 cancer-sample-specific somatic mutations (S1 Table), of which 42,380 (50.1%) were located in protein-coding (coding) regions, and the remaining 42,191 (49.9%) were located in noncoding regions (Table 1, Fig 1A). The noncoding regions included (i) 5'UTRs, (ii) ~100 bp fragments of 3'UTRs adjacent to coding sequences (3'UTRs), (iii) exon-adjacent ~100 bp fragments of introns (introns), and (iv) sequences other than those classified above (i-iii), mostly intergenic sequences located upstream and downstream of the first and last gene exons (intergenic regions). The average coverage of the mutated positions was 169x and was slightly higher in coding (195x) than in noncoding regions (142x), whereas the average fraction of reads mapping to alternative alleles was 0.35 (0.33 in coding and 0.40 in noncoding regions).

The average mutation rate in the targeted regions was 52.8 mutations/Mb (ranging from 0.1 to 287.5), which, although slightly lower than that observed before in BCC [15,34], is still higher than that in any other tested cancer type. Although somewhat counterintuitive, the lower mutation burden in our study than in other BCC studies [15,34] may result from the much higher sequencing coverage in our study, which gave us much higher statistical power to filter out the fraction of false-positive mutations. The lower mutation burden in our study may also be explained by the identification in our cohort of two samples with an extremely low mutational burden (<0.2 mutations/Mb). Most of the identified mutations were single-nucleotide substitutions (79,960 (94.5%), predominantly C>T transitions), followed by double
substitutions (3,128 (3.7%), predominantly CC>TT transitions) and short (<4 nt) indels (1,483 (1.8%)) (Table 1, Fig 1B). The higher frequency of indels in noncoding regions most likely results from the excess of low complexity sequences, which cause polymerase slippage.

To estimate the fraction of false-positive mutations, we resequenced (with Sanger sequencing) 52 mutations representing different types of alterations, including 39 substitutions and 13 indels (S2 Table). The analysis confirmed 51/52 of the mutations, indicating a very low (2%) fraction of false-positive mutations. The fraction may be even lower, as the only unconfirmed mutation (double substitution CC>TT in MYCN) was present in a low fraction of reads (7%), which is generally beyond the sensitivity of Sanger sequencing.

**Mutational signatures**

The analysis of mutational signatures showed that most of the samples were predominantly associated with signature 7 (average signature contribution (SC) = 0.7) and to a lesser extent with signature 11 (average SC = 0.2) (Fig 1C-D). Both signatures consist predominantly of C>T substitutions but differ in the sequence context of the substitutions. Signature 7 is associated with UV irradiation exposure and commonly occurs in melanoma and head and neck cancer. A hallmark of signature 7 is the frequent occurrence of double CC>TT substitutions resulting from UV radiation-induced pyrimidine dimers. Signature 11 was previously found in melanoma and glioblastoma multiforme, often in patients treated with the alkylating agent temozolomide, which is also used in BCC therapy. Only one sample (BCC22) showed a stronger association with signature 11 (SC = 0.6) than signature 7 (SC = 0.3). None of the analyzed samples showed an association with signatures 1, 2, 5, and 13, which are frequent in most cancer types. This may indicate that the deamination of 5-methylcytosine (5meC) predominantly induced by AID/APOBEC cytidine deaminases (attributed to the abovementioned signatures) does not play a role in the pathogenesis of BCC.
The comparison of the nodular and superficial BCC samples showed no substantial difference in terms of mutation burden or mutation types, with the exception of the contribution to mutational signature 7, which was higher for the nodular than superficial samples (Fig 1E), consistent with the higher UV radiation exposure of nodular BCCs.

**Hotspot mutations**

As recurrent mutations may be indicators of the cancer-related function of the mutated genes, we first looked for hotspots defined as genomic positions mutated in at least 3 samples (>10% of the cohort). In total, we identified 43 hotspots, including 23 hotspots in coding and 20 hotspots in noncoding regions (8 in 5'UTRs, 1 in 3'UTRs, and 11 in introns) (S3 Table). Of the coding hotspots, 16 resulted in missense mutations, and 7 were synonymous substitutions. Although the functionality of individual synonymous mutations cannot be unequivocally ruled out [35–37], the majority of these hotspot mutations resulted from randomly occurring neutral alterations, so we did not analyze them further. As shown in S3 Table, some of the hotspots were located in genes annotated in the COSMIC Cancer Gene Census (CGC) database and/or in genes playing a role in cancer or skin function.

**Hotspot mutations in coding regions**

Of the coding mutations (S3 Table), the most commonly identified in our study (in 5 samples) was the c.1292C>T (Ser431Phe) substitution, located at chr14:103,131,144 in the Sec6 domain of TNFAIP2, which encodes a multifunctional protein playing a role in angiogenesis, inflammation, cell migration and invasion, cytoskeleton remodeling, and cell membrane protrusion formation [38–41]. Nonetheless, TNFAIP2 is not well-recognized in cancer, and the hotspot or other mutations in the gene have not been reported before. Another coding hotspot, mutated in 3 samples with the c.655C>T (Pro219Ser) substitution, was located at chr7:148,827,237 in EZH2; EZH2 encodes an essential subunit (methyltransferase) of
polycomb repressive complex 2 (PRC2), which plays a role in histone methylation and gene silencing [42]. *EZH2* is a well-known oncogene associated with a more aggressive form and poorer prognosis of many cancers, including melanoma, squamous cell carcinoma (SCC), and BCC, with demonstrated increased expression in SCC (compared to normal skin and SCC precursor actinic keratosis (AK)) [43] and aggressive BCC [44]. Both gain- and loss-of-function mutations in *EZH2* have often been found in myeloid leukemias and lymphomas but are not common in solid tumors. Contrary to the previously detected mutations clustering mostly in the catalytic SET domain (Bödör et al. 2013; Donaldson-Collier et al. 2019), the hotspot detected here was located in the N-terminal (NT) part of the protein, which, among other areas, is responsible for interaction with histones [47]. Whether the mutations may affect the interaction warrants further investigation. To the best of our knowledge, this mutation hotspot has not been observed in any cancer, including BCC.

An additional interesting coding hotspot (mutated in 3 samples) was located at chr15:40,382,906-40,382,907. The hotspot was mutated with either the c.71C>T substitution or the c.71_72delinsTT double substitution (note that double substitutions are annotated as deletion/insertion (delins) variants according to HGVS nomenclature), both resulting in the Ser24Phe missense mutation affecting the NT part of the KNSTRN protein (also known as small kinetochore-associated protein (SKAP)), which plays a role in maintaining chromatid cohesion and proper chromatid separation during anaphase [48]. *KNSTRN* mutations (predominately the Ser24Phe hotspot mutation) were first detected in 19% of SCCs and 13% of AKs [49]. Subsequent analysis of The Cancer Genome Atlas (TCGA) datasets showed that the *KNSTRN* mutations also occur in 5% of melanoma samples but are rare in other cancers. Later, *KNSTRN* mutations were also identified in 2% [15] and 10% [50] of BCCs. These findings together with this study confirm that *KNSTRN* mutations are specific to UV radiation-related skin cancers. Consistent with the role of KNSTRN, it was shown that *KNSTRN*
mutations in SCC affect proper chromosome separation and are associated with increased chromosome instability, expressed as a fraction of the genome with copy number alterations (CNAs) [49]. Although there was a similar number of tested samples, the association of the KNSTRN mutations with CNAs was not confirmed in BCC, neither in a study by Jaju et al. [50] nor in our study (S1 Fig). It is worth noting that it was also shown that KNSTRN plays a role in UV radiation-induced apoptosis [51]; however, the effect of the mutations on avoidance of apoptosis by BCC cells or any other cancer cells has not yet been tested.

**Hotspot mutations in noncoding regions**

The most frequently mutated hotspot of all the hotspots detected in the study hotspot (mutated in 8 samples) was located at chr11:64,270,066-64,270,067 in the 3'UTR of **BAD** and has never been reported before. The hotspot encompasses 4 different substitutions (c.*142C>A, c.*142C>T, c.*142_*143delinsTT and c.*143C>T; S3 Table and Fig 2A), located 142 or 143 nucleotides (nt) downstream of the stop codon. The protein encoded by the gene is a member of the BCL-2 family, which plays a role in the positive regulation of cell apoptosis. The gene is commonly implicated in many cancers [52,53]; however, to the best of our knowledge, this hotspot has not been reported before in any cancer.

Next, another novel noncoding hotspot mutated in 5 samples located at chr16:72,008,760-72,008,761 in the 5'UTR of **DHODH** was identified. The hotspot encompasses two different substitutions, c.-5G>A and c.-5_-4delinsAA, affecting the Kozak sequence (S3 Table and Fig 2B). **DHODH** is not well studied in cancer, but it has recently been demonstrated that it plays an important role in the carcinogenesis of SCC and other UV radiation-induced skin cancers [54,55].

Another mutated noncoding hotspot from our study worth mentioning was found in 4 samples with the c.-77C>T substitution and was located at chr7:56,106,490 in the 5'UTR of **CHCHD2**, also known as **MNRR1** (S3 Table and Fig 2C). The analysis of the entire **CHCHD2** 5'UTR
showed one more recurrent (in 2 samples) substitution, c.-134G>A, located at chr7:56,106,547, resulting in a total of 6 mutations in the 5'UTR in 6 samples. Interestingly, frequent mutations in the hotspot in the 5'UTR of CHCHD2 were previously reported in melanoma [56].

Finally, we identified a hotspot located at chr1:153,990,763 in the 5'UTR of RPS27 (encoding a ribosomal protein component of the 40S subunit) that was mutated in 3 samples with the c.-34C>T substitution. Mutations in the promoter/5'UTR of RPS27 (including the hotspot mutation) have been identified before in ~10% of melanoma samples [56,57] but have never been reported in BCC or other skin cancers. Subsequent in vitro functional studies showed that the RPS27 5'UTR hotspot mutation decreases RPS27 mRNA levels and that decreased levels of RPS27 are associated with worse prognosis of melanoma patients and drug (vemurafenib and palbociclib) sensitivity of melanoma cells [58].

**Computational analysis of the identified noncoding hotspots and comparison with external datasets**

To further characterize three noncoding hotspot mutations, two not previously reported in BAD and DHODH and one in CHCHD2 previously reported in melanoma [56], we analyzed their potential impact with a number of computational tools and investigated their incidence in other cancers using external datasets of a large cohort (>10,000 samples) of TCGA samples, representing 33 different human cancer types (including 469 skin cutaneous melanoma (SKCM) samples but not including BCC or SCC samples). Note that the list and the standard abbreviations of all TCGA cancer types are in S4 Table.

In total, in the TCGA samples, we identified 28 mutations in the BAD 3'UTR (Fig 2A). The mutations were found predominantly in SKCM samples(15 mutations in 12 (2.6%) SKCM samples), including 4 mutations in the hotspot (residues c.*142C and c.*143C) identified in BCC, and 6 c.*77C>T mutations, constituting an additional hotspot in the 3'UTR, not occurring in BCC. In other cancers, 3'UTR mutations were very rare (Fig 2A). In contrast with the
mutation frequency in the 3'UTR, mutations in other parts of the gene, including the coding region (n=26, predominantly missense or synonymous), were rare (not exceeding 1% in any cancer) and randomly distributed between different cancer types (excluding SKCM). The exclusiveness of the SKCM and BCC mutations in the 3'UTR vs. other parts of the gene (enrichment compared to other cancer types; Fisher's exact test; p<0.0001 and p=0.0005, respectively) precludes an accidental occurrence of the mutations, solely as a result of some region- and/or mutagenesis-related mechanisms and argues for the cancer-driven selection of the 3'UTR mutations in BCC and SKCM (and likely also in other UV irradiation-related cancers).

Next, with the use of TargetScan, we identified 3 miRNAs (miR-7515, miR-3202, and miR-6125) whose predicted targets (seed-interacting sequences) were disrupted by hotspot mutations (Fig 2A). However, as (i) none of these targets has been validated by any means (miRTarBase; [59]), (ii) none of these miRNAs have been confidently validated (via miRBase or miRGeneDB), and (iii) none of these miRNAs have been found to have expression levels detectable/confirmed in any of the TCGA cancers, it is very unlikely that any of the identified targets are functional. Additionally, the occurrence of SKCM mutations in different positions across the BAD 3'UTR argues against the possibility that the driving force of the mutations is a disruption of a particular miRNA target. Some clue for the functionality of the BCC hotspot may be its location in the 5' arm of the ~40 bp long stable hairpin RNA structure motif (dG=-39.6 Kcal/mol), which is destabilized (by ~2 Kcal/mol) by the hotspot mutations (Fig 2A).

The analysis of TCGA data showed no mutation in the BCC hotspot or any other mutation in the DHODH 5'UTR in any of the TCGA cancer types, even though different mutations (n=81) were identified in other parts of the gene, including 75 mutations in the coding region (Fig 2B). The other mutations, however, were randomly distributed along the gene sequence and between different cancer types, and only two of the coding mutations were deleterious (frameshift).
mutations. This result indicates that the DHODH 5'UTR hotspot mutations are BCC-specific mutations, and the absence of these mutations in other UV radiation-related cancers makes it unlikely that the frequent occurrence of the mutations in BCC is solely due to a random effect of UV irradiation. The 5’UTR of DHODH is very short (21 bp). Although hotspot mutations occurred in the Kozak sequence, which is important for the initiation of translation, neither wild-type nor mutant alleles affected the consensus Kozak sequence nucleotides (at positions -4 and -5); therefore, the ATGpr [60]), and NetStart 1.0 (GedersenAG [61] tools predicted the mutations to have a minor effect on the effectiveness of translation under standard conditions. However, this result does not exclude an effect of the mutations under specific conditions, such as hypoxia, UV exposure, or cancer.

The analysis of RNA secondary structure showed that the hotspot mutations slightly modified (decreased the stability of) a small hairpin motif predicted to be formed by an RNA sequence directly flanking the hotspot (Fig 2B). The mutation may also destabilize the potential long-range interaction of the sequence flanking the mutations with the sequence located ~200 nt downstream. Analysis of the 5’UTR sequence [62] showed that the double substitution (GG>AA) at the hotspot creates a consensus binding site for the NFAT1 transcription factor (Fig 2B), which is expressed in many tissues, including sun-exposed and non-sun-exposed skin (GTExPortal; GTEx Consortium Science 2020), and implicated in many cancers, including melanoma [63,64].

In total, in TCGA data, we identified 63 mutations in the CHCHD2 5'UTR (Fig 2C). The mutations were found predominantly in SKCM samples (40 mutations in 39 (8.5%) samples), including 29 c.-77C>T mutations and 3 c.-134G>A mutations, located in the hotspot positions identified in BCC. Additionally, we identified 4 samples with the c.-74C>T mutation, constituting an additional hotspot in the 5'UTR. Only 5 SKCM mutations were located outside the 5'UTR, 4 in the CDS (2 missense and 2 synonymous) and 1 in the 3'UTR (one mutation).
(Fig 2C). In other cancers, there were rare 5'UTR mutations, including 4 mutations in HNSC and UCEC, 3 mutations in BRCA, and 12 mutations in other cancers. Three of these mutations coincided with the c.-77 hotspot. The positions of BCC/SKCM hotspot mutations seem to be nonrandom because they were all located in and all disrupted two distinct GABP-alpha transcription factor binding sites (mapped with the use of MotifMap [65]) (Fig 2C).

**Frequently mutated genes**

Next, we looked at the overall frequency of mutations in the genes, separately analyzing mutations in coding regions, 5'UTRs, 3'UTRs, and introns (defined in Materials and Methods; listed in S5 Table). Although they were not considered frequently mutated, in this section, we also report genes with any mutations in a coding region if they were detected in a pathway of a recurrently mutated gene. In the analysis of frequently mutated regions, we focused mostly on genes functionally related to cancer (annotated with CGC and a manual literature search) and genes playing a role in skin function.

**Genes frequently mutated in coding regions**

In total, we identified 606 genes frequently mutated in coding regions. The most frequently mutated was *PTCH1*, with a total of 24 mutations in 20 BCC samples, including 5 missense, 4 splice-site, and 15 deleterious (nonsense or frameshift) mutations (Fig 3A). Mutation c.3450-1G>A located upstream of exon 21 was one of the splice-site mutations and was also observed in another study [14], which suggests its recurrence in BCC. We tested and confirmed the exon-skipping effect of the mutation with the use of exon-junction PCR and Sanger sequencing analysis (Fig 3A). The other genes from the hedgehog pathway recurrently mutated in our cohort were *GLI2*, which was mutated in 5 samples, and *SMO*, which was mutated in 4 samples (S4 Fig and Fig 4). The combined frequency of *SMO* and *GLI2* mutations was much lower in samples with (4/20; 20%) than in those without (4/7; 57%) *PTCH1* mutations, which suggests
mutual exclusiveness of these mutations (Fig 4). Altogether, 24 (88%) samples had mutations in genes involved in the hedgehog pathway. Other frequently mutated cancer-related genes were TP53 (7 missense, 8 deleterious, and one splice-site mutation in 13 samples) (Fig 3B); MYCN (8 missense mutations in 8 samples), NOTCH1 (8 missense and 2 deleterious mutations in 8 samples), NOTCH2 (3 missense, 3 deleterious, and 2 splice-site mutations in 7 samples), NOTCH3 (6 missense mutations in 5 samples; note that the NOTCH mutations colocalized with the regions of the loss-of-function mutations identified in other solid tumors, e.g., in SCCs [66], LATS1 (5 missense and one deleterious mutation in 5 samples), and ARID1A (5 missense mutations in 5 samples) (Fig 4 and S2 Fig). The mutations in the abovementioned genes are generally consistent with mutations observed before in BCC [14,15]. Additionally, we identified very frequent mutations (18 missense and 1 deleterious) in PTPRD (Fig 3C), a tumor suppressor frequently mutated in many cancers, including melanoma and cutaneous SCC [67–71], in 13 samples, but these have never been reported as frequently mutated in BCC. Interestingly, in addition to mutations in MYCN, we also noticed recurrent (although not frequent) mutations in three other genes in the MYC/MTOR regulatory network, i.e., MTOR, Dyrk3, and AMBRA1 (Fig 4), which have not been reported as mutated in BCC. The MTOR missense/activating mutations identified in other cancers are considered biomarkers for therapy with mTOR pathway inhibitors [72]).

Finally, we found a high frequency of mutations in the FLG (15 mutations in 10 samples) and FLG2 (9 mutations in 9 samples) genes (Fig 3D-E and Fig 4), encoding profilaggrin and filaggrin-like proteins, precursors of filaggrin. Filaggrin is an important component of the stratum corneum of the epidermis that plays a role in maintaining epithelial homeostasis and barrier functions [73] and is a substrate for trans-urocanic acid (UCA) and pyrrolidone carboxylic acid (PCA), which are suggested to serve as a natural UV radiation barrier [74]. Although frequent mutations in the FLG/FLG2 genes have been previously observed in other
cancers, the mutations were usually considered random (passenger). Here, however, we observed a relatively high proportion of deleterious nonsense mutations, altogether occurring in 6 samples. Additionally, the analysis of the entire cohort of TCGA samples showed that the frequency of the FLG/FLG2 mutations observed in our study in BCC substantially exceeds the frequencies of the mutations in other cancers, including melanoma (the next most frequently mutated cancer) (S3 Fig).

**Genes frequently mutated in noncoding regions**

Among the 11 genes frequently mutated in the 5'UTR (S5 Table) there were DHODH and CHCHD2 with the hotspot mutations, as described in the previous paragraph. Of interest may also be SPHK2, with 4 dispersed mutations in 4 samples, whose function as both a proapoptotic gene suppressing cell growth and an oncogene promoting cell proliferation has been proposed [75–79]. SPHK2 also had mutations in its coding region (Fig 4).

Among the 11 genes frequently mutated in the 3'UTR (S5 Table), in addition to BAD, which was described in the previous paragraph, we also identified 8 mutations in the 3'UTR of SMIM27 (also annotated as IncRNA TOPORS-AS1); the overexpression of SMIM27 was found to be associated with favorable outcomes in breast cancer [80].

Finally, we identified 289 genes (15 annotated in CGC) frequently mutated in introns (S5 Table). Interestingly, among the genes was PTCH1, which, in addition to 4 splice-site mutations (mentioned above), also had other 4 intronic mutations (in total, 8 intronic mutations). Other genes with frequent mutations in introns included PTPRD (14 mutations in 9 samples), which also frequently had mutations in the coding region; NOTCH2 (6 mutations, including 2 splice-site mutations in 6 samples), which also frequently had mutations in the coding region; ERBB4 (6 mutations in 6 samples), a well-known oncogene playing a role in many cancers (reviewed in [81]); and DROSHA (5 mutations in 5 samples), which encodes a
core enzyme (nuclease) of the miRNA processing pathway and has been shown to be upregulated in BCC [82].

**Driver genes in BCC (OncodriveFML analysis)**

To further investigate the mutations/mutated genes, we used OncodriveFML, which allows the prediction of the cancer driver potential of both coding and noncoding regionsgenes based on functional mutation (FM) bias [83]. As shown in Fig 5A-C and S6 Table, we identified 14 potential cancer driver genes based on mutations in coding regions (CDS-drivers), a disproportionately high number of 36 potential cancer driver genes based on mutations in 5'UTRs (5'UTR-drivers), and 7 potential cancer driver genes based on mutations in 3'UTRs (3'UTR-drivers). No potential cancer driver gene was identified based on the mutations in introns.

In addition to 4 CDS-drivers (*PTCH1*, *TP53*, *TGFB1I1*, and *CARD6*) also identified as frequently mutated, it is worth noting *RORA*, recently shown to play an important role in restraining allergic skin inflammation [84]. Other interesting genes were *PRDM9* and *ZNF281*, both of which play a role in DNA repair and have been shown to be responsible for frequent mutations in cancer [85,86]. None of these genes were previously implicated or identified as frequently mutated in BCC.

Among the 5'UTR-drivers, 6 were also identified as frequently mutated: *DHODH*, *CHCHD2*, and *SPHK2* (described above), as well as *POLR2M*, *NPC1*, and *NEL2*. Additionally, it is worth noting *IKBKB* (mutated in 3 samples but not reported before as mutated in BCC) shown to act as a tumor suppressor in nonmelanoma skin cancers and noncancerous skin lesions; it was also shown that deletions of the gene lead to skin inflammation, hair follicle disruption, hyperplasia, and SCC development [87–90].
Among 3'UTR-drivers, two genes (mentioned above), i.e., BAD (the most significant 3'UTR-driver) and SMIM27 were also identified as frequently mutated. Additionally, it is worth mentioning the transcription factor gene POU3F2 (mutated in 3 samples), that plays a role in the invasiveness and metastasis of melanoma, and is controlled by miR-211 [91,92] and miR-107 [93]. Although the mutations were not located in the predicted miR-107 and miR-211 binding sites, they may affect the structure of the 3'UTR and thus indirectly change accessibility to these or other miRNA targets.

Analysis of copy number alterations

As somatic CNAs have not been extensively studied in BCC, in the next step, we performed analysis of both chromosome arm-level and focal CNAs (with GISTIC2; [94]). At the chromosome arm level, we detected a significant recurring deletion of chr9q (q=1.4x10^-6; occurring in 9 samples), involving PTCH1 (Fig 4 and Fig 6), and a significant recurring amplification of chr9p (q=0.05; occurring in 5 samples), involving a region with CD274 (also known as PDL1, encoding PD-L1), CD273 (also known as PDL2, encoding PD-L2), and JAK2 (Fig 4 and Fig 6). Although the loss of chr9q has been frequently observed in BCC (reported as loss-of-heterozygosity of PTCH1), gain of chr9p has been reported only in one case of rare metastatic BCC [95]. To validate the chromosome 9 CNAs, we developed a multiplex ligation-dependent probe amplification (MLPA) assay with probes covering the entire chromosome 9 but especially focusing on the region containing PTCH1 (chr9q22.32) and the region harboring PDL1, PDL2, and JAK2 (chr9p24.1) (Fig 6). The MLPA analysis confirmed CNAs in all tested samples as detected by GISTIC2, and examples are shown in Fig 6.

CNA analysis also showed 54 regions of significant focal deletions, including 27 regions containing skin/cancer-related genes, and 56 significant amplifications, including 20 encompassing skin/cancer-related genes (Fig 6 and S7 Table). The elements involved in the
most significant focal deletions were \textit{CDK11A} \textup{(}chr1p36.33; \textup{q}=2.4x10^{-5}; occurring in \textup{6 samples}), whose loss induces skin carcinogenesis \cite{96}; the \textit{LCE} cluster \textup{(}chr1q21.3; \textup{q}=2.4x10^{-6}; occurring in \textup{4 samples}), including genes such as \textit{LCE2} and \textit{LCE3}, which play a role in maintaining skin barrier function and whose deletion has been associated with psoriasis \cite{97}; and the \textit{HLA-D} cluster \textup{(}HLA-\textit{DP}, -\textit{DQ}, and -\textit{DR}, chr6p21.32; \textup{q}=2x10^{-4}; occurring in \textup{3 samples}), encoding components of major histocompatibility complex (MHC) class II molecules, whose increased expression has been associated with increased cancer immunogenicity and better prognosis in BCC, SCC and melanoma \cite{98–104}. The skin/cancer-related genes in the most significant focally amplified regions worth mentioning are \textit{STIM2} \textup{(}chr4p15.2; \textup{q}=0.16; occurring in \textup{2 samples}) \cite{105}, \textit{KLRB1/CD161} \textup{(}chr12p13.31; \textup{q}=0.007; occurring in \textup{2 samples}) \cite{106,107}, and \textit{SPTLC3} \textup{(}chr20p12.1 \textup{q}=0.23; occurring in \textup{2 samples}) \cite{108}. 
DISCUSSION

In this study, we detected thousands of mutations in BCC samples, many of which were clustered in specific genes/regions or hotspots located in both coding and noncoding regions. Despite the small size of our dataset, our results are in line with those of previous genomic analyses of coding mutations in BCC [14,15], which confirms the reliability of our study. We believe that our results may give valuable insights related to general characteristics of mutations such as mutational burden or mutational signatures and in terms of genes identified as recurrently mutated in coding regions.

Moreover, we extended our analysis to noncoding parts of the genes, which altogether were responsible for ~50% of the mutations identified by standard WES approach. Variants in such areas have usually been ignored in previous BCC genetic studies. Many of the identified noncoding hotspots were located in sequences of genes functionally related to cancer or more specifically to UV radiation-related skin cancers. Some of them were reported before in melanoma or identified by us in melanoma TCGA samples, the cancer type most intensively studied in terms of mutations in noncoding regions [109,110]. Below, we briefly describe the cancer-related role of the three most interesting genes with hotspot mutations in noncoding regions, i.e., BAD, DHODH, and CHCHD2. Interestingly, all these genes have functions related to mitochondrial activity.

Of all the hotspots detected in our study, the most frequently mutated was the hotspot located in the 3'UTR of BAD. This hotspot had several different mutations affecting 2 nucleotide positions (142 and 143 nt downstream of the stop codon). Due to these mutations, BAD was also classified as being highly mutated in the 3'UTR and as the top most significant potential cancer driver. Consistently, the hotspot and several other positions in the 3'UTR are frequently mutated in melanoma but not in other cancers. BAD belongs to the BCL-2 family, consisting of both proapoptotic and antiapoptotic proteins. It promotes cell death by inducing
mitochondrial outer membrane permeabilization (MOMP), allowing the release of cytochrome c, and by antagonizing (dimerizing with) antiapoptotic BCL-2 proteins [111,112]. On the other hand, phosphorylated BAD may also have antiapoptotic properties, e.g., promoting the survival of melanocytes [113,114]. Other functions of BAD include regulation of mitochondrial metabolism (regulation of voltage-dependent anion channels and metabolite passage through the outer mitochondrial membrane) and dynamics (regulation of shape changes) [115–121]. Although BAD has not been previously implicated in skin cancers, loss or downregulation of other proapoptotic members of the BCL-2 family, i.e., BAX and PUMA, has been shown to promote the development of BCC, SCC, and cutaneous melanoma [122,123]. Therefore, a similar effect may be induced by mutations causing more efficient downregulation of BAD.

*CHCHD2* is a gene with frequent mutations in the 5'UTR, the hotspot mutation c.-77C>T and the recurrent mutation c.-134G>A (77 and 134 upstream of the start codon). Based on the 5'UTR mutations, *CHCHD2* was classified as a high priority cancer driver. We showed that the *CHCHD2* 5'UTR (predominantly the hotspot position) was also frequently mutated (8%) in the SKCM TCGA samples, which also showed the additional recurrent mutation c.-74C>T. The 5'UTR mutations were also found in whole-genome sequenced Australian melanoma samples [56]. The role of the gene has not been intensively studied in cancer, but it was shown that under hypoxic conditions, CHCHD2 is translocated from the mitochondrial intermembrane space to the nucleus, where it binds an oxygen-responsive element in the promoter of *cytochrome oxidase 4I2 (COX4I2)*, encoding a subunit of complex IV of the electron transport chain, and increases its expression. Consequently, *CHCHD2* knockdown downregulates *COX4I2* and decreases cell oxygen consumption [124]. It was also shown that CHCHD2 is a negative regulator of mitochondria-mediated apoptosis [125]. Liu et al. showed that CHCHD2 interacts with antiapoptotic BCL-XL (from the BCL-2 family), which leads to inhibition of proapoptotic BAX and consequently decreases MOMP and apoptosis. In addition, it was shown
that CHCHD2 dysregulates multiple genes that play a role in cell migration and cancer metastasis and that its expression is higher in cell lines derived from more aggressive breast tumors [126]. Consistent with the function of CHCHD2 related to mitochondrial metabolism, we found that all BCC/SKCM hotspot/recurrent mutations coincided with and impaired two distinct binding sites of GABP-alpha. As GABP-alpha is known to be a transcription factor involved in the regulation of cellular energy metabolism and cell cycle regulation [127], this finding might hint at a functional role of the mutations in cancer. Of note, germline missense mutations in CHCHD2 are associated with autosomal dominant Parkinson's disease [128].

DHODH is a gene that showed frequent mutations in the Kozak sequence of the 5'UTR, with hotspot mutations encompassing two different substitutions, c.-5G>A and c.-5_4delinsAA (4 and 5 nt upstream of the start codon). Based on the identified mutations, DHODH was classified as a candidate cancer driver. The analysis of the entire TCGA cohort (~10K samples from 33 cancer types) showed that no other cancer had mutations in the hotspot or the 5'UTR, indicating that the mutations were BCC-specific. Although DHODH 5'UTR mutations have never been reported before in any cancer, it was shown very recently that DHODH plays a key role in the carcinogenesis of SCC and other UV radiation-induced skin cancers and facilitates the development of precancerous skin lesions [54,55]. Hosseini et al. showed that the DHODH protein level and enzymatic activity are markedly upregulated in irradiated skin and that an increased level of DHODH sensitizes the skin to UV irradiation-induced damage. It was also shown that DHODH is upregulated in melanoma, in which DHODH inhibition leads to a marked decrease in tumor growth both in vitro and in mouse xenograft studies [129]. DHODH inactivation inhibits cell proliferation and induces cell cycle arrest at the S phase in BCL-2 (pro-apoptotic) deficient melanoma cells (Liu et al. 2017). DHODH is embedded in the inner mitochondrial membrane, and its canonical role is in the oxidation of dihydroorotate to orotate, an important step in de novo pyrimidine synthesis (which is important in replication and DNA
repair). However, a side product of the pathway, ubiquinol (QH2), is a source of electrons in the electron transport chain, and DHODH also plays a role in alternative (glucose-independent) respiration (utilizing amino acids as an energy source) [54,55,130], facilitating cancer development in hypoxic conditions. In addition, it was found that in esophageal SCC, elevated DHODH levels promote cell proliferation by stabilizing β-catenin [131]. The functional effects of the mutations may result from alteration of the Kozak sequence but also the creation of an NFAT1 transcription factor binding site, which is not present in the wild-type sequence. NFAT1 is a widely distributed isoform of the NFAT family of transcription factors and is expressed in tumor cells and the tumor microenvironment [132]. The constitutive activation and overexpression of NFAT1 in many cancer types promote the transcription of genes that are crucial for cancer development and progression, including COX2, MMP7, MMP9, and MDM2 [133,134].

It is worth noting that noncoding mutations in only the promoters of TERT and DPH3 were previously analyzed in BCC [27,135,136]; TERT and DPH3 are known to be mutated in many cancers, including melanoma [109,110]. However, as our WES design generally did not cover promoter regions, we did not look for or detect these mutations.

Additionally, the whole-genome CNA analysis allowed us to detect two highly significant chromosome-level CNAs. In addition to the expected deletion of chr9q, consistent with the loss of heterozygosity of PTCH1, we also detected frequent duplication/amplification of chr9p, encompassing the PDL1 and PDL2 genes (which encode the two immune checkpoint proteins PD-L1 and PD-L2, the overexpression of which enables cancer cells to evade the host immune system). Copy number gains of PDL1 have been observed only in one case of metastatic BCC [95]. The patient, who was otherwise resistant to vismodegib and sonidegib, demonstrated a dramatic response to nivolumab (an anti-PD-1 antibody blocking the PD-1/PD-L1 interaction), which strongly suggested that the copy number gain may be a biomarker of sensitivity to anti-
PD-1/PD-L1 checkpoint treatments [95]. It was also shown in an independent study that some patients (up to ~40%) with advanced BCC (not tested for PDL1 amplification) respond to pembrolizumab (another anti-PD-1 antibody) [95,137]. Therefore, assessment of copy number gains of the PDL1/PDL2 region may help to rationalize such treatment; however, further study with a larger group of samples is required.

Finally, we would like to note the apparent limitations of the study. As it was intended to be a preliminary evaluation of noncoding mutations in BCC, we analyzed only a small number of samples, and as such, we limited the characterization of the identified variants to computational analyses.

In summary, in this study utilizing WES BCC data, we revealed not only mutations in coding regions of previously known BCC-related genes but also frequent mutations in noncoding regions of cancer-related genes, some of which may be strong candidates for new BCC drivers. Although the functional role of the individual identified genes/mutations requires further experimental interrogations, our results provide a strong basis for further analyses of noncoding variants in BCC and other cancer types.
MATERIALS AND METHODS

Sample collection and DNA preparation

A total of 27 pairs of tissue (tumor and normal adjacent healthy skin) were collected from the Department of Plastic Surgery, St. Josef Hospital, Catholic Clinics of the Ruhr Peninsula, Essen, Germany. While excising the BCC tissues with cold steel under local anesthesia, 4-mm punch biopsies were taken from the center of the tumor and from nonlesional epithelial skin (as normal, intraindividual controls). These samples were immediately placed in RNAlater (Qiagen, Hilden, Germany) and stored at −80 °C. Tissue homogenization was performed with stainless steel beads of 5 mm (Qiagen) and TissueLyser LT (Qiagen). DNA was extracted with an AllPrep DNA/RNA/miRNA Universal Kit (Qiagen) according to the manufacturer’s protocol. All samples were quantified using a NanoDrop One (Thermo Scientific, Waltham, USA) and Qubit fluorometer 3.0 (Invitrogen) (Qubit dsDNA HS Assay (Life Technologies, Carlsbad, USA)), and DNA size and quality were tested using gel electrophoresis.

Exome sequencing and data processing

The library was prepared with 200 ng of high-quality DNA using the SureSelectXT Library Prep Kit (Agilent). A SureSelectXT Human All Exon V6 kit (Agilent) was used for exome capture. Sequencing was performed on an Illumina NovaSeq 6000 (San Diego, USA), generating 2x 100 bp paired-end reads. Library preparation, exome enrichment, and sequencing were performed at CeGaT, Tuebingen, Germany. Demultiplexing of the sequencing reads was performed with Illumina bcl2fastq (2.19). Adapters were trimmed with Skewer (version 0.2.2) [138]. The Phred score was given with Illumina standard Phred encoding (offset +33). For each sample, two FASTQ files corresponding to forward and reverse reads were obtained. Paired-end reads were aligned to hg38 using BWA. PCR duplicates were marked and removed with
the Picard package. Indel realignments with known sites and base quality score recalibration were performed with GATK version 4.1.2.0. SAM to BAM conversion was done using SAMtools. Somatic single-nucleotide variants were called with MuTect2 (version 4.1.0.0. with the use of the tumor-normal mode). To distinguish somatic from germline mutations, we relied on the gnomAD database (version 2.1.1) as a public germline mutation source. We also generated a panel of normals (PON) file, which contained a set of variants that were detected by MuTect2 when run on a set of normal samples (but these were not germline mutations). From the VCF (variant call format) files, we extracted somatic mutation calls with PASS and panel of normals annotation. We also added information about the localization of mutations in gene subregions (CDS, 5′UTR, 3′UTR, or introns) by use of an in-house Python script. From the list of somatic mutations, we additionally removed those that did not fulfill the following criteria: (i) at least five alternative allele-supporting reads in a tumor sample; (ii) frequency of alternative allele-supporting reads in a tumor sample of at least 0.05; and (iii) frequency of alternative allele-supporting reads in the tumor sample at least 5× higher than that in the corresponding normal sample.

Validation of mutations

A panel of 51 mutations detected by WES was validated by Sanger sequencing of the appropriate PCR fragments amplified with primers shown in S8 Table. All fragments were sequenced in two directions with the BigDye v3.1 kit (Applied Biosystems, Foster City, CA, USA), and the sequencing reactions were separated with capillary electrophoresis (POP7 polymer; ABI Prism 3130xl apparatus; Applied Biosystems, Foster City, CA, USA) according to the standard manufacturer’s recommendations.
Mutational signature analysis

To analyze mutational signatures, we used the web application Mutational Signatures in Cancer (MuSiCa; http://bioinfo.ciberehd.org/GPtoCRC/en/tools.html [139]), allowing the visualization of the somatic mutational profile of each analyzed sample and estimation of the contribution values of the predefined mutational signatures ([140]; Catalogue Of Somatic Mutations In Cancer, COSMIC 2020). Samples BCC14 and BCC21 were excluded from the signature analysis due to an insufficient number of mutations.

Identification of hotspots, frequently mutated genes, and cancer drivers

We defined genomic positions mutated in at least 3 samples as hotspots. Mutations occurring in directly adjacent nucleotides were merged into one hotspot.

We defined genes with nonsynonymous mutations in a coding region in at least 5 samples, with mutations in a 5'UTR, in at least 4 samples, with mutations in a 3'UTR in at least 4 samples, and with mutations in introns (up to 40 nt from exon/intron boundaries) in at least 5 samples as frequently mutated. From the analysis, we excluded genes known to be commonly hypermutated with passenger mutations as a result of the increased background mutation rate but not related to cancer, listed in [141]. To distinguish synonymous from nonsynonymous mutations, we used the SnpEff - genetic variant annotation and functional effect prediction toolbox [142], available on the Subio platform (Subio, Inc., Kagoshima, Japan, http://www.subio.jp). We also considered splice-site mutations located in introns up to +/-2 nt from exons as coding region mutations.
OncodriveFML [83] was run using the CADD score (hg38, version 1.6). The signature method was set as a complement, the statistical method was set to “amean”, and indels were included in the analysis using a max method (max_consecutive was set to 7 as default).

Copy number analysis

To identify chromosome arm-level and focal regions that were significantly amplified or deleted, we used GISTIC2 [94] with the following parameters: threshold for copy number amplifications and deletions, 0.2; confidence level to calculate the region containing a driver, 0.9; broad-level analysis; and the arm peel method to reduce noise.

To validate CNAs involving chromosome 9, i.e., chr9p duplications/amplifications (affecting JAK2, PDL1/CD274, and PDL2/CD273) and chr9q deletions (affecting PTCH1), we designed and generated an MLPA assay covering the entire chromosome 9. In total, the assay consisted of 20 probes, including (i) 7 probes distributed over the chr9p (n=5) and chr9q (n=2) arms, 2 probes located in or in close proximity to JAK2, PDL2, PDL1, and PTCH1 (in total 8 gene-specific probes), and 5 control probes (located on different chromosomes outside of chromosome 9 and regions of known cancer-related genes). The sequences and detailed characteristics of all probes as well as their exact positions are shown in S9 Table and are schematically depicted in Fig 6C.

The MLPA probes and the probe-set layout were designed according to a previously proposed and well-validated strategy [143,144].

TCGA analysis

To compare the mutations recurring in BCC with mutations in other cancers, we used WES-generated somatic mutation datasets of 10,369 samples representing 33 cancer types generated
and deposited in the TCGA repository (http://cancergenome.nih.gov). The full names and abbreviations of all TCGA cancer types are shown in S4 Table. Somatic mutations were identified against matched normal samples with the use of the standard TCGA pipeline (including the Mutect2, Muse, Varscan, and SomaticSnipper algorithms). We extracted somatic mutation calls (with PASS annotation only) localized in the annotated exons of BAD, DHODH, CHCHD2, FLG, and FLG2 (exon sequences were extended by 2 nt to enable identification of intronic splice-site mutations). The extraction was performed as described in our earlier study [145] with a set of in-house Python scripts available at (https://github.com/martynaut/mirnaome_somatic_mutations).

Mutations visualization

All mutations were annotated according to HGVS nomenclature (at the transcript and protein levels), and the effects of mutations were defined using the Ensembl Variant Effect Predictor (VEP) tool. For visualization of mutations on gene maps, we used ProteinPaint from St. Jude Children’s Research Hospital – PeCan Data Portal [146]. The protein domains visualized on gene maps were positioned according to UniProt data [147]. The comutation plot showing frequently mutated genes was created with the use of the Python library CoMut [148].

Analysis of RNA regulatory motifs

Target predictions were performed with the TargetScan Custom (release 5.2) web tool [149]. The secondary RNA structures were predicted using mfold software [150] with default parameters. RNA sequence/structure functional motifs and transcription factor binding sites were analyzed with the RegRNA 2.0 [151] and MotifMap [65] web tools.
Statistics

Specific statistical tests are indicated in the text, and a $p$-value $<0.05$ was considered significant. If necessary, $p$-values were corrected for multiple tests with the Benjamini-Hochberg procedure.

FUNDING

This work was supported by research grants from the Polish National Science Centre [2016/22/A/NZ2/00184] (to P.K.).
Author Contributions

Conceptualization: Michael Sand, Piotr Kozlowski

Data curation: Paulina Nawrocka, Paulina Galka-Marciniak, Martyna Urbanek-Trzeciak

Formal analysis: Paulina Nawrocka, Paulina Galka-Marciniak, Martyna Urbanek-Trzeciak, Natalia Szostak, Piotr Kozlowski

Funding acquisition: Piotr Kozlowski

Investigation: Paulina Nawrocka, Paulina Galka-Marciniak, Martyna Urbanek-Trzeciak, Ilamathi M., Piotr Kozlowski

Methodology: Paulina Nawrocka, Paulina Galka-Marciniak, Martyna Urbanek-Trzeciak, Piotr Kozlowski

Project administration: Paulina Nawrocka, Piotr Kozlowski

Resources: Laura Susok, Michael Sand

Software: Paulina Nawrocka, Paulina Galka-Marciniak, Martyna Urbanek-Trzeciak, Natalia Szostak, Anna Philips

Supervision: Piotr Kozlowski

Validation: Paulina Nawrocka

Visualization: Paulina Nawrocka, Paulina Galka-Marciniak, Piotr Kozlowski

Writing - original draft: Paulina Nawrocka, Piotr Kozlowski

Writing - review & editing: Paulina Nawrocka, Paulina Galka-Marciniak, Martyna Urbanek-Trzeciak, Ilamathi M., Natalia Szostak, Anna Philips, Laura Susok, Michael Sand, Piotr Kozlowski
REFERENCES


119. Roy SS, Madesh M, Davies E, Antonsson B, Danial N, Hajnóczky G. Bad targets the permeability transition pore independent of Bax or Bak to switch between Ca2+-dependent cell survival and death. Mol Cell. 2009;33: 377–388. doi:10.1016/j.molcel.2009.01.018


# TABLE

Table 1. Summary of somatic mutation distribution and mutation types in BCC.

<table>
<thead>
<tr>
<th>Genomic regions</th>
<th>No. (%) of mutations</th>
<th>Average coverage of mutation positions</th>
<th>Average alternative allele fraction</th>
<th>No. (%) of substitutions</th>
<th>No. (%) of double substitutions</th>
<th>No. (%) of short indels</th>
</tr>
</thead>
<tbody>
<tr>
<td>all mutations</td>
<td>84,571 (100)</td>
<td>169</td>
<td>0.35</td>
<td>79,960 (94.5)</td>
<td>3128 (3.7)</td>
<td>1483 (1.8)</td>
</tr>
<tr>
<td>coding</td>
<td>42,380 (50.1)</td>
<td>195</td>
<td>0.33</td>
<td>40,503 (95.6)</td>
<td>1697 (4.0)</td>
<td>180 (0.4)</td>
</tr>
<tr>
<td>noncoding</td>
<td>42,191 (49.9)</td>
<td>142</td>
<td>0.37</td>
<td>39,457 (93.5)</td>
<td>1431 (3.4)</td>
<td>1303 (3.1)</td>
</tr>
<tr>
<td>introns</td>
<td>32,805 (38.8)</td>
<td>136</td>
<td>0.38</td>
<td>30,655 (93.4)</td>
<td>1076 (3.3)</td>
<td>1074 (3.3)</td>
</tr>
<tr>
<td>3’UTR</td>
<td>2,926 (3.5)</td>
<td>154</td>
<td>0.39</td>
<td>2,721 (93.0)</td>
<td>100 (3.4)</td>
<td>105 (3.6)</td>
</tr>
<tr>
<td>5’UTR</td>
<td>2,832 (3.3)</td>
<td>163</td>
<td>0.37</td>
<td>2,650 (93.6)</td>
<td>125 (4.4)</td>
<td>57 (2.0)</td>
</tr>
<tr>
<td>intergenic</td>
<td>3,628 (4.3)</td>
<td>167</td>
<td>0.35</td>
<td>3,431 (94.6)</td>
<td>130 (3.6)</td>
<td>67 (1.8)</td>
</tr>
</tbody>
</table>
Fig 1. Mutation distribution, mutational signatures, and comparison of superficial and nodular BCC subtypes. A) Frequency of mutations in particular gene/genomic regions. B) Frequency of mutation types. C) Heatmap showing the contribution of the mutational signatures (rows) to the analyzed BCC samples (columns). A higher color intensity indicates a higher contribution (as indicated on the scale bar). D) Representative mutation distribution plots of samples with a high association with signature 7 (sample BCC25) and signature 11 (sample BCC22). E) Comparison of nodular and superficial BCC samples in terms of (from the left) mutational load, signature 7 and signature 11 contributions, frequency of indels, and frequency of double substitutions.
Fig 2. Distribution of mutations in the selected genes with the identified mutation hotspots in noncoding areas.

A-C) Maps of the BAD, DHODH, and CHCHD2 genes, with the exon structure and protein functional domains indicated. Mutations are visualized in the form of lollipop plots along with the gene maps, and the size of a mutation symbol (circle) is proportional to the number of mutations. Mutations identified in BCC (red) are shown above and mutations identified in SKCM (blue) and other TCGA cancers (gray) are shown below the maps. The inset below each map shows the detailed sequence context of the hotspot mutations, along with CADD score graphs, indicating the functional relevance of particular positions and other sequence characteristics (i.e., (in A) predicted miRNA target sites, (in B) the Kozak consensus sequence and NFAT1 transcription factor binding sites (BSs) created by the hotspot mutation, and (in C) the GABP-alpha transcription factor BSs disrupted by the hotspot mutations). The additional insets in A and B show computationally predicted RNA secondary structures generated from RNA sequences directly flanking the hotspots.
Fig 3. Distribution of the identified mutations in the genes with frequent mutations in the coding sequence. A)-E) Maps of the PTCH1, TP53, PTPRD, FLG, and FLG2 genes. Mutations are visualized in the form of lollipop plots along with gene maps; the size of a mutation symbol (circle) is proportional to the number of mutations, and the color indicates the type of mutation (as shown in the legend). Additionally, the inset in A) shows the Sanger sequencing reads depicting the effect of the splice-site mutation c.3450-1G>A on exon 21 skipping.
Fig 4. Comutation plot summarizing the somatic alterations in the BCC samples. Columns correspond to the samples, and rows correspond to the selected genes. The color of the mutation presence symbols corresponds to the mutation type, as indicated in the legend on the right. The bar plots above and on the left indicate the mutational burden and the fraction of samples with mutations in particular genes, respectively. The nodular and superficial samples are indicated by color.

Fig 5. Identification of potential cancer drivers with the use of OncodriveFML. The quantile-quantile (QQ) plots show the distribution of expected (x-axis) and observed (y-axis) p-values corresponding to FM bias calculated (with CADD score) separately for mutations in A) coding regions, B) 5’UTRs, and C) 3’UTRs. The green and red colors indicate genes defined as significant (q<0.025) and highly significant (q<0.01), respectively, according to OncodriveFML recommendation.
Fig 6. CNA analysis of the BCC samples. A) GISTIC-estimated q-values for deletions (left, blue) and amplifications (right, red) are plotted along with chromosome positions (vertically). The green line indicates the recommended significance threshold, q=0.25. The selected significantly deleted and amplified regions/genes are indicated on the graphs. B) Representative MLPA results (bar plots), showing samples with chromosome 9 CNAs, i.e., chr9q deletion and chr9p amplification, vs. a sample (at the top) with the wild-type (WT) copy number genotype. Each bar plot depicts relative copy number values (y-axis) of the probes specific for regions along chromosome 9 and an average (with standard deviation error bar) signal of control probes (x-axis). C) Schematic depictions of the localization of the probes on chromosome 9 and in genes of interest.
Supporting information

S1 Fig. Concurrence of KNSTRN mutations and CNAs.
The number of amplifications (red dots), deletions (blue dots), and total CNAs (orange dots) in samples with and without the KNSTRN mutations (red font - hotspot mutation; green font - other mutation).
S2 Fig. Distribution of the identified mutations in the selected genes frequently mutated in coding sequence. A)-F) Maps of MYCN, NOTCH1, NOTCH2, NOTCH3, LAT51, and ARID1A, respectively.
S3 Fig. Frequency of mutations in the *FLG* and *FLG2* genes in BCC and other cancer types. Bar graphs showing the frequency of mutations in A) *FLG*, and B) *FLG2* in BCC (current study, green bar) and TCGA cancer types.

S4 Fig. Frequency of activating (red) and deactivating (blue) mutations in genes of the hedgehog pathway.
S1 Table. List of somatic mutations in BCC.

(XLSX)

S2 Table. List of somatic mutations validated by Sanger sequencing.

(XLSX)

S3 Table. List of hotspot mutations.

(XLSX)

S4 Table. TCGA abbreviations and full cancer/project names.

(XLSX)

S5 Table. List of frequently mutated genes.

(XLSX)

S6 Table. List of driver genes identified by the OncodriveFML analysis.

(XLSX)

S7 Table. List of significant focal CNAs.

(XLSX)

S8 Table. List of the primers used for PCR.

(XLSX)

S9 Table. List of the probes used for the MLPA assay.

(XLSX)