Whole-exome sequencing identifies a somatic missense mutation of NBN in clear cell sarcoma of the salivary gland

LEI ZHANG1*, ZHEN JIA2*, FENGBIAO MAO3,4*, YUEYI SHI1, RONG FA BU1 and BAORONG ZHANG2

1Department of Stomatology, Chinese PLA General Hospital, Beijing 100853; 2Department of Stomatology, Beijing Aviation General Hospital, Beijing 100012; 3Beijing Institutes of Life Science, Chinese Academy of Sciences, Beijing 100101; 4University of the Chinese Academy of Sciences, Beijing 100049, P.R. China

Received December 12, 2015; Accepted January 15, 2016

DOI: 10.3892/or.2016.4738

Abstract. Clear cell sarcoma (CCS) is a rare, low-grade carcinoma commonly located in the distal extremities of young adults involving tendons and aponeuroses. CCS is characterized by its poor prognosis due to late diagnosis, multiple local recurrence, propensity to late metastases, and a high rate of tumor-related mortality. The genetic cause for CCS is thought to be EWSR1 gene translocation. However, CCS lacking a translocation may have other, as yet uncharacterized, genetic mutations that can cause the same pathological effect. A combination of whole-exome sequencing and Sanger sequencing of cancer tissue and venous blood from a patient diagnosed with CCS of the salivary gland revealed a somatic missense mutation, c.1061C>T (p.P354L), in exon 9 of the Nibrin gene (NBN). This somatic missense mutation led to the conversion of proline to leucine (p.P354L), resulting in deleterious effects for the NBN protein. Multiple-sequence alignments showed that codon 354, where the mutation (c.1061C>T) occurs, is located within a phylogenetically conserved region. In conclusion, we here report a somatic missense mutation c.1061C>T (p.P354L) in the NBN gene in a patient with CCS lacking an EWSR1-ATF1 fusion. Our findings broaden the genotypic spectrum of CCS and provide new molecular insight that should prove useful in the future clinical genetic diagnosis of CCS.

Introduction

Clear cell sarcoma (CCS), regarded as malignant melanoma of soft parts, is a rare aggressive tumor that accounts for less than 1% of all soft tissue sarcomas (1,2). CCS is characterized by its poor prognosis due to late diagnosis, multiple local recurrence, propensity to late metastases, and a high rate of tumor-related mortality (3,4). As one of the few sarcomas with a high propensity for lymph node metastases (5), CCS is a locally aggressive neoplasm with a high rate of recurrence and metastasis (more than 50%) (2,6,7). The 5-year disease-specific survival rates have been reported to be approximately 50-67%, but these values are not representative of long-term survival since many patients develop lung and bone metastases more than five years after initial resection (7,8). This tumor is observed most frequently in young adults and predominantly affects the soft tissues of the distal extremities, with the majority being deep seated and involving tendons and aponeuroses (4,9). Occasionally, it can arise in visceral organs including the gastrointestinal tract (10). It is rarely seen in the head or neck, with for example only 1.2% of the approximately 500 reported cases of CCS currently involving the head or neck (1).

Clear cell sarcoma was first described by Dr Franz M. Enzinger in 1965 (11). CCS is characterized by a nested or fascicular growth pattern of spindled and epithelioid cells with clear or lightly eosinophilic cytoplasm surrounded by fibrous septa (12). The tumor cells have elongated oval nuclei with prominent nucleoli and occasional nuclear pseudo-inclusions (13). Multinucleated giant cells are identified in more than half of the reported cases (7). Uniquely among primary soft tissue tumors, pre-melanosomes are present in almost all cases of CCS, detectable by electron microscopy (14). As a result, immunohistochemistry (IHC)-based tests of CCS cells are almost always positive for the melanoma markers S-100, HMB45, MelanA, and microphthalmia transcription factor (MITF), although melanin staining is not always observed (12,15). Adverse prognostic factors for CCS identified to date include large tumor size and any microscopic tumor necrosis. Surgery is the mainstay of treatment for this high grade sarcoma, with chemotherapy having little effect. Although the melanocytic differentiation of CCS is indisputable, its precise lineage remains unclear. Thus, CCS maintains the status of a unique yet enigmatic clinicopathological entity (4).

As a rare type of soft tissue sarcoma, CCS exhibits morphological, immunohistochemical and ultrastructural similarity with malignant melanoma (5). CCSs share many
features with malignant melanoma, including expression of melanoma markers (16). However, in contrast to most melanomas, CCSs lack BRAF mutations (9). In addition, molecular analysis has revealed that CCSs are distinct tumors; they present the specific t(12;22)(q13q12) translocation that results in the chimeric gene EWSR1-ATF1, which is not observed in melanomas (5). The genetic cause for CCS is considered to be this defining gene translocation. Previous cytogenetic studies have established the specificity of the recurrent t(12;22)(q13,q12) translocation, resulting in an EWSR1-ATF1 fusion for CCS (17). However, EWSR1-ATF1 fusion derived from cytogenetic rearrangements is characteristic but not entirely unique for CCS, as similar fusion genes are also present in angiomatoid fibrous histiocytoma (18). Generally, detection of this fusion gene and the absence of BRAF gene mutations easily distinguish CCS from cutaneous melanoma (4).

Having said that, it must be noted that not all CCSs present with EWSR1 rearrangements. According to a previous study, approximately 70% of CCS cases harbored a rearrangement (EWSR1-ATF1 or EWSR1-CREB1) in the EWSR1 locus with a mean of 81.6% positive cells/sample (range, 60-95%) (19). The more prevalent fusion event in CCS, EWSR1-ATF1, also occurs in both hyalinizing clear cell carcinoma (20) and angiomatoid fibrous histiocytoma (18). However, the exons involved are different, with most CCS tumors involving EWSR1 exons 7, 8, or 10 fused to ATF1 exons 4, 5, or 7 (21), but with HCC tumors harboring EWSR1 exon 11 fused in-frame to exon 3 of ATF1 (20). Therefore, the specific EWSR1-ATF1 fusion in CCS can typically be used to distinguish CCS from its mimics such as spindle cell melanoma, spindle cell squamous carcinoma, cutaneous leiomyosarcoma and atypical fibroxanthoma, as well as from other tumors with melanocytic differentiation (12).

To date, the pathogenesis of CCS lacking an EWSR1 rearrangement remains poorly characterized. No somatic mutations involved in CCS have been identified, but single nucleotide variants are known to play a significant role in tumorigenesis (22). In the present study, we identified a somatic missense mutation c.1061C>T (p.P354L) in exon 9 of the Nibrin (NBN) gene in a patient with CCS of the salivary gland via a combination of exome sequencing and Sanger sequencing. It is known that although mutations of NBN do not play a major role in predisposition to melanoma of the skin, alterations in this gene may contribute to the risk for breast cancer (23-25).

NBN is a protein associated with the repair of double-strand breaks (DSBs), which cause serious damage to genomes. NBN is a 754 amino acid protein known to be a member of the NBS1/hMre11/RAD50 double-strand DNA break repair complex (referred to as MRN) (26). This complex recognizes DNA damage and rapidly relocates to DSB sites, forming nuclear foci. NBN also has a role in the regulation of MRN protein complex activity, including involvement in end-processing of both physiological and mutagenic DNA DSBs. The mutations within exons 6–10 of the NBN gene in patients suffering Nijmegen breakage syndrome (NBS) result in a truncated protein (27). Patients with NBS are predisposed to cancers. This predisposition to cancer may be linked to the DSBs that occur during the development of lymphoid cells. Moreover, mutations of NBN have been reported to be associated with many types of cancers, including gastrointestinal lymphoma (28), childhood acute leukaemia (29), glioblastomas (30), and breast cancer (25). In this study, we report the first known somatic mutation of NBN that was found to be involved in CCS of the salivary gland, and provide molecular insight into future clinical genetic diagnosis for CCS by broadening the genotypic spectrum of CCS.

Materials and methods

Subject and statement of ethics. The subject was a 20-year-old Chinese male that presented with a gradual mass of duration 3-4 months located on the left parotid gland. The mass was diagnosed as clear cell sarcoma (CCS) of the salivary gland by incisional biopsy.

The experimental methods used in this study were carried out in accordance with the relevant guidelines and regulations. This study was approved by the Ethics Committee of the General Hospital of the Chinese People's Liberation Army. Relevant informed consent was obtained from all participants, including the CCS patient and the healthy control subjects.

All sequencing data for the tumor and venous blood samples have been deposited in the Sequence Read Archive (SRA, http://www.ncbi.nlm.nih.gov/sra/) with accession nos. BioSample SAMN03384326 and SAMN03384334, respectively, in BioProject SRP055838.

Exome sequencing. The tumor and venous blood from the Chinese male CCS patient was selected for exome sequencing. Exome sequencing was carried out using an Agilent SureSelect Human All Exon v5.0 (51M) kit, according to the instructions from Illumina's TruSeq Exome Enrichment Guide (SureSelectXT Target Enrichment System for Illumina Paired-End Sequencing Library, Agilent). Genomic DNA libraries were prepared according to the manufacturer's instructions (Illumina Inc., USA). Briefly, 3 μg of genomic DNA was randomly fragmented into pieces of 100-500 bp in size using a Diagenode Bioruptor® system (Diagenode). DNA fragments between 150 and 250 bp were recovered by gel extraction. An end repair and size selection procedure was then performed with T4 DNA polymerase and Klenow polymerase cleavage 3’. An ‘A’ base was added to the 3’ end of the fragments using Klenow 3’ to 5’ exo minus. The DNA fragments were then ligated to the Illumina multi-PE-adaptor. The adapter-ligated templates were purified using Agencourt AMPure SPRI beads and amplified by four-cycle ligation-mediated polymerase chain reaction (LM-PCR), under the following PCR conditions: 2 min at 94°C, four cycles of 10 sec at 94°C, 30 sec at 62°C, and 30 sec at 72°C, and then 5 min at 72°C. The LM-PCR products were hybridized to the Agilent Oligo pool for 24 h at 65°C for enrichment. The hybridized fragments were bound to streptavidin beads and non-hybridized fragments were washed out. Captured LM-PCR products were amplified by PCR (2 min at 98°C; 10-12 cycles for 10 sec at 98°C, 30 sec at 60°C, 30 sec at 72°C, followed by 5 min at 72°C). The magnitude of enrichment of the samples was estimated with an Agilent 2100 bioanalyzer. The captured library was then sequenced on Illumina HiSeq 2000 analyzers with 126 cycles/read, to generate paired-end reads and 8 bp of index tag (following the manufacturer's standard sequencing instructions).
Read mapping and variant analysis. Image analysis and base calling were performed with Illumina Basecaller program (v1.8). Indexed primers were used for data fidelity surveillance. The sequence reads were aligned to the human genome reference obtained from the UCSC database (http://genome.ucsc.edu/), version hg19 (GRCh37), using the SOAP aligner program (v2.21). Single nucleotide polymorphisms (SNPs) were called using SOAPsnp (v1.03) with the default parameters, after the duplicated reads (produced mainly in the PCR step) had been removed using Picard (v1.63) (31). Short insertions or deletions (InDels) altering coding sequence or splicing sites were identified by GATK (v1.4-33-g051b450) through realignment analysis of insertions and deletions, quality recalculation, and InDels calling (UnifiedGenotyper in GATK). We then filtered candidate SNPs with the following criteria: SNP quality, ≥20; sequencing depth, ≥10; the estimated copy number, ≤2; and the distance between two SNPs, >5. Subsequently, we used VarScan2 v2.3.7 (http://varscan.sourceforge.net/) to detect the somatic mutations in the exome data from tumor-normal pairs. We applied varElect (http://varelect.genecards.org/) to select mutations associated with the function of salivary glands. The effect of candidate mutations to protein features were predicted with the GERP++ program (May 22, 2011). Potential rejected substitutions were evaluated by PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/) and SIFT prediction (http://sift.jcvi.org/).

PCR and Sanger sequencing. Following the analysis of the exome sequencing data, Sanger sequencing was performed to verify the detected genetic variants. Primers flanking the mutation area of NBN (NM_002485) were designed based on the reference genomic sequences of the human genome from NCBI GenBank, and synthesized in Shanghai, China, by Thermo Scientific. The sequence of the forward primer was 5’-CCCTACCTCATTGGCTTTGTG-3’, and that of the reverse primer was 5’-TATCACGGTCCCTGCTTCC-3’. All PCR amplification was carried out using an Applied Biosystems Life Technologies (ABI) 9700 thermal cycler. PCR products were directly sequenced on an ABI PRISM 3730 automated sequencer (Applied Biosystems Life Technologies). Sequence comparisons and analyses were performed using the Jalview program (v2.8.2; http://www.jalview.org/).

Results

Clinical phenotype. A 20-year-old male presented with a gradual mass of duration of 3-4 months on the left parotid gland. He occasionally suffered from a short period of needle pain. On palpation, a sub-mucosal mass measuring 3x2x2 cm in size was confirmed. The mass was non-tender and varied in consistency from soft to firm. On general examination, the patient appeared apparently healthy with neither submandibular nor cervical lymph node metastases. The mass was diagnosed as CCS of the salivary gland by incisional biopsy (Fig. 1A-D). Histologically, the tumor had typical features including a combination of cords and nests of clear and eosinophilic cells with haphazard arrangement and intersecting dense collagen bands is present in clearing for CCS of the salivary gland (x200 magnification, H&E). Immunohistochemical staining for HMB45 is positive in tumor cells (x100 magnification, immunohistochemistry). Immunohistochemical staining for CK is negative in tumor cells (x200 magnification, immunohistochemistry).
tumor cells actually had a pale eosinophilic cytoplasm rather than a clear cytoplasm, or they may have had a mixture of both (Fig. 1A). Focal squamous differentiation was also noted occasionally (Fig. 1C). There was a tendency for the cells in the center of the mass to be admixed with, or to be surrounded by, a hyalinized basement membrane-like material (Fig. 1D). Tumor cells at the periphery of the mass had a greater tendency for nest formation and for wide infiltration without a desmoplastic response or stromal deposition (Fig. 1A). Moreover, immunohistochemical staining revealed that the tumors were HMB45-positive (Fig. 1C) and CK-negative (Fig. 1D). Based on these observations and lines of evidence, we diagnosed this tumor as CCS of the salivary gland.

**Mutation analysis.** We performed exome sequencing of the tumor tissue from the CCS patient. We generated 19.28 and 7.95 billion bases of 125-bp paired-end read sequences from the tumor and from the venous blood samples, respectively. Billion bases [19.09 (99%) and 7.88 (99%)] passed the quality assessment, and 19.01 (98.61%) and 7.91 (99.46%) billion bases aligned to the human reference for the tumor and venous blood samples, respectively. For tumor tissue, 11.39 billion bases (59.9%) mapped to the targeted regions with a mean coverage of 80.47 X 21,077 genetic variants, including 9,295 non-synonymous variants, were identified in either the coding regions or the splice sites. For venous blood, 4.61 billion bases (58.3%) mapped to the targeted regions.
with a mean coverage of 63.42 X; 19,930 genetic variants, including 8,404 non-synonymous variants, were identified in either the coding regions or the splice sites. A prioritization scheme was applied to identify the pathogenic mutation in the tumor tissue, using methods similar to those reported in recent studies (32,33). We excluded the known variants that had been identified in the dbSNP141, 1000 genomes, HapMap, and ESP-6500 datasets. After filtering the variants displayed in venous blood, we obtained 542 candidate somatic mutations in the tumor tissue.

Subsequently, we used varElect software to select mutations associated with the function of the salivary gland and acquired four candidate causal mutations. A somatic mutation, c.1061C>T (p.P354L), was eventually identified in exon 9 of the NBN gene in the tumor and this was validated using Sanger sequencing. This mutation results in a missense variant. The same mutation was absent in peritumoral tissue of the patient (Fig. 2) and venous blood samples from 30 ethnically matched normal control individuals. This mutation was also absent in the dbSNP141, 1000 genomes, HapMap, and ESP-6500 datasets. To evaluate whether or not there were EWSR1-ATF1 or EWSR1-CREB1 gene translocations, RT-PCR (reverse transcription polymerase chain reaction) was performed to validate the fusion site; the negative result was thus confirmed.

**Bioinformatic analysis of NBN mutations in CCS.** We obtained sequences for NBN family proteins using the BLAST tools of the NCBI databases and performed multiple-sequence alignments using Jalview in various animal species, including Homo sapiens, Pongo abelii, Pan paniscus, Ailuropoda melanoleuca, Cricetulus griseus and Bos taurus (Fig. 3). The p.Pro354Leu variant was found to be located in a highly conserved region of the NBN protein, suggesting its likely structural and functional importance. This mutation was predicted to affect the protein features and be rejected.

Figure 4. A snapshot of histone modification from the ENCODE project in the locus around the NBN mutation, as viewed with the Integrative Genomics Viewer tool.
Figure 5. A snapshot of transcriptional factors of from the ENCODE project in the locus around the NBN mutation as viewed with the Integrative Genomics Viewer tool.

Figure 6. Network analysis of NBN using the dSysMap ('Disease-mutations Systemic Mapping') program.
substitutions predicted by GERP++ with a score 3.07. SIFT prediction indicated a deleterious effect for this mutation, with a score of 0. In addition, PolyPhen-2 prediction also suggested that this mutation probably conferred a damaging effect, with a confident score of 0.998.

GO annotations related to NBN include damaged DNA binding and transcription factor binding. However, it remains unclear which histone modifications regulate the expression of NBN and it is not yet clear which transcriptional factors are associated with the mutation locus of NBN. Using integrated analysis of the data for histone modifications deposited in the Encyclopedia of DNA Elements (ENCODEx) project (34), we found that histone modifications at the NBN locus and the c.1061C>T mutant variant include H3K36me3, H3K27ac, and H3K4me1/2/3 in multiple cell lines such as GM12878, H1-hESC and K562 (Fig. 4). As these histone modifications are canonically active molecule markers (35), we speculated that the c.1061C>T variant may influence histone modification and affect the expression of NBN, although it must be noted that the histone modification data are from cell lines rather than from CCS tumors. Indeed, according to our PCR validation of expression analysis, the expression of NBN was decreased in the CCS tumor compared with its expression level in peritumoral tissue. Similarly, several transcriptional factors deposited in the ENCODE project were found to be enriched by ChIP-seq in the locus where the c.1061C>T variant include H3K36me3, H3K27ac, and H3K4me1/2/3 in multiple cell lines such as GM12878, H1-hESC and K562 (Fig. 4). As these histone modifications are canonically active molecule markers (35), we speculated that the c.1061C>T variant may influence histone modification and affect the expression of NBN, although it must be noted that the histone modification data are from cell lines rather than from CCS tumors. Indeed, according to our PCR validation of expression analysis, the expression of NBN was decreased in the CCS tumor compared with its expression level in peritumoral tissue. Similarly, several transcriptional factors deposited in the ENCODE project were found to be enriched by ChIP-seq in the locus where the c.1061C>T variant occurred in the NBN gene for cell line GM12878 (Fig. 5). These transcriptional factors were identified as E2F4, SPT20, BRCA1, NRF1, TR4, NFKB and WHIP; these may function to regulate the transcription of the NBN gene.

Discussion

Clear cell sarcoma (CCS) of soft tissue, formerly referred to as malignant melanoma of soft parts, is a neoplasm with poor prognosis that primarily affects young adults between the ages of 20 and 40 years (36). The tumor has a high propensity for lymph node metastasis and local recurrence. CCS typically involves tendons and aponeuroses. Primary CCS of the salivary glands is exceedingly rare. Only a few cases of primary CCS arising in the ulna, metatarsals, ribs, radius, sacrum, humerus (37) and jejunum (10) have been reported, and to the best of our knowledge, our case of CCS arising in the salivary glands is only the second such study to date.

CCS is a translocation-associated sarcoma. In chromosomal translocations, the pieces of two chromosomes are swapped; this can result in an abnormal fusion of genes. Most cases of CCS harbor a fusion of EWSR1/ATF1 resulting from translocation (21). The genetic cause for CCS is thought to be its defining gene translocation. However, 30% of cases of CCS have no observed EWSR1 translocation (10). A CCS lacking a translocation may have other, as yet uncharacterized, genetic mutations that can cause the same pathological effect.

Here, we report a somatic missense mutation c.1061C>T (p.P354L) in the NBN gene of a Chinese patient with CCS that did not harbor the typically expected EWSR1-ATF1 fusion. The pathogenesis of the ~30% of CCS cases lacking an EWSR1 rearrangement has always been a mystery (19). The EWSR1-ATF1 fusion is one of the characteristics used to distinguish CCS from cutaneous melanoma (4). In contrast to most melanomas, CCS cells lack BRAF mutations (9) and show immunohistochemical positive staining for HMB45 and negative staining for CK (12,15). In our case study, the immunohistochemical results were consistent with the diagnostic criteria of CCS. Moreover, we observed a combination of cords and nests of clear and eosinophilic cells in a hyalinized background, which is a typical feature of CCS in histological diagnoses. Therefore, the finding of this novel NBN mutation provides a novel genotypic feature for the clinical genetic diagnosis of CCS.

This mutation in NBN was located within a phylogenetically conserved region, suggesting that it has a role of structural and/or functional importance. Previous studies have indicated that mutations in NBN are associated with Nijmegen breakage syndrome, an autosomal recessive chromosomal instability syndrome characterized by microcephaly, growth retardation, immunodeficiency and cancer predisposition. The NBN protein is a member of the MRE11/RAD50 double-strand break (DSB) repair complex, which consists of 5 proteins including NBN, MDC1, FANC2D, RR1M2B and MRE11A (Fig. 6). This NBN protein is thought to be involved in DNA DSB repair and DNA damage-induced checkpoint activation (38). Cancers associated with NBN include gastrointestinal lymphoma (28), childhood acute leukemia (29), glioblastomas (30), breast cancer (25), prostate cancer (39) and ovarian cancer (40). Given these known associations, the mechanism of carcinogenesis for CCS may be related to or share similar pathway(s) associated with other NBN-related cancers. This intriguing supposition will require further investigation. In conclusion, our findings broaden the genotypic spectrum of CCS and provide new molecular insight that should prove valuable for the future clinical genetic diagnosis for CCS.

Acknowledgements

We thank all subjects for their participation. We also thank the scientists responsible for various web sources, including the 1000 Genomes Project (http://www.1000genomes.org), dbSNP (http://www.ncbi.nlm.nih.gov/SNP), ESP-6500 (http://evs.gs.washington.edu/EVS/), HapMap (http://hapmap.ncbi.nlm.nih.gov/) and the various DNA damage prediction tools including SIFT (http://sift.jcvi.org) and PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2), among others.

References


