

Genetic Analysis of Sinonasal Undifferentiated Carcinoma Discovers Recurrent SWI/SNF Alterations and a novel PGAP3-SRPK1 Fusion Gene

Molly Heft Neal

University of Michigan–Ann Arbor

Andrew Birkeland

University of Michigan–Ann Arbor

Apurva Bhangale

University of Michigan–Ann Arbor

Jingyi Zhai

University of Michigan–Ann Arbor

Aditi Kulkarni

University of Michigan–Ann Arbor

Susan Foltin

University of Michigan–Ann Arbor

Brittany Jewell

University of Michigan–Ann Arbor

Megan Ludwig

University of Michigan–Ann Arbor

Lisa Pinatti

University of Michigan–Ann Arbor

Hui Jiang

University of Michigan–Ann Arbor

Jonathan McHugh

University of Michigan–Ann Arbor

Lawence Marentette

University of Michigan–Ann Arbor

Erin McKean

University of Michigan–Ann Arbor

J Brenner (✉ chadbren@umich.edu)

University of Michigan–Ann Arbor

Keywords: SNUC, SWI/SNF, SMARCA

Posted Date: December 28th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-115402/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Sinonasal Undifferentiated Carcinoma (SNUC) is a rare and aggressive skull base tumor with poor survival and limited treatment options. To date, targeted sequencing studies have identified *IDH2* and *SMARCB1* as potential driver alterations, but the molecular alterations found in *SMARCB1* wild type tumors are unknown.

Methods: We evaluate survival outcomes in a cohort of 46 SNUC patients treated at an NCI designated cancer center and identify clinical and disease variables associated with survival on Kaplan-Meier and Cox multivariate survival analysis. We perform exome sequencing to characterize a series of *SMARCB1* wild type tumors and cell line including identification of high confidence mutations, copy number alterations, microsatellite instability, and fusions. Knockdown studies using siRNA was utilized for validation of a novel *PGAP3-SRPK1* gene fusion.

Results: We discover recurrent aberrations to the SWI/SNF and FAT gene families. We also validate a novel *PGAP3-SRPK1* gene fusion in the SNUC cell line, and show that knockdown of the fusion is negatively associated with EGFR, E2F and MYC signaling.

Conclusion: Collectively, these data demonstrate recurrent alterations in the SWI/SNF and FAT gene families and discover a novel fusion gene (*PGAP3-SRPK1*). These data aim to improve understanding of possible driver mutations and guide future therapeutic strategies for this disease.

Background

Sinonasal Undifferentiated Carcinoma (SNUC) is a highly aggressive disease involving the anterior skull base, nasal cavity and paranasal sinuses. It is a rare tumor, with only a few hundred cases in the literature (1). Patients usually present at an advanced stage, and have poor outcomes (2, 3), with two-year overall survival rates as low as 25% in some cohorts (1, 4–9). Validated prognostic factors are limited to traditional clinical variables (overall stage, high grade, and poor differentiation), and no additional data on possible informative biomarkers is currently in clinical use (10). Current treatment modalities including surgery, radiation, and systemic chemotherapy alone or in combination with radiation (CRT) have poor outcomes and carry significant toxicity to patients (11–13). A recent study reveals improved survival with chemoselection paradigms, with five-year disease free survival rates of 59% in the total cohort and rates as high as 81% in responders (14). However, despite these promising results, patients who did not show initial response to induction chemotherapy had a 0% and 39% five-year DSS when treated with CRT and surgery +/- CRT respectively. These results indicate the urgent need for novel therapeutics particularly for this subset of patients with aggressive disease. Importantly, there have been no novel or targeted agents introduced for SNUC treatment since its initial identification, which is partially due to a limited investigation into the underlying genetics defining SNUC pathogenesis.

To date, only a few case reports describing mutations associated with disease pathogenesis have been published. The most commonly reported mutations include *IDH2* and *SMARCB1* which have been

identified in small case series via traditional sequencing approaches or targeted sequencing panels (15–17). There have been additional case reports of potentially actionable mutations in isolated SNUCs including *ERBB2* and *FGFR1* (18, 19), but previous efforts have been limited in their scope of sequencing (4) and currently there have been no comprehensive whole exome or genome sequencing studies performed on SNUCs.

As such, this rare, devastating disease has limited treatment options currently available and characterizing genomic profiles of SNUCs may have significant benefit for the future development of rational therapeutic strategies. By understanding the genomic architecture behind this disease process, we may also begin to identify prognostic biomarkers that help identify the patients that fail current treatment paradigms. Here, we provide survival data from 46 patients treated at our tertiary referral center and report the first whole exome sequences profiling the mutational landscape of SNUCs.

Materials And Methods

Patient Population

A single-institution retrospective case series informed by a prospectively maintained database of patients with SNUC was performed. The study was approved by the University of Michigan Institutional Review Board (HUM00080561). Patients with a history of sinonasal undifferentiated carcinoma treated at the University of Michigan were included in the clinical dataset (n = 46). Inclusion criteria for genomic, copy number and transcriptome analysis is as follows: 1) Patients with sinonasal undifferentiated carcinoma as confirmed by our board-certified pathologist (J.B.M.); 2) Blocks maintained in the University of Michigan pathology archive; 3) Sufficient DNA or RNA yield for next generation sequencing. Additionally, a prospective patient was consented to our University of Michigan IRB-approved MiOT0seq precision medicine program (HUM00085888) as described (20). In total, there were 5 patients who met inclusion criteria for analysis, and demographics are shown in **Supplemental Table 1**.

Survival Analysis Statistics:

Survival was calculated using Kaplan-Meier analysis and outcomes were compared using Log-rank analysis. Multivariate cox regression analysis was performed using Backward Wald method with an inclusion of variables with p-values < 0.1. Statistical analysis was performed using SPSS v26 (IBM, Armonk, NY). Kaplan-Meier curves were created using Prism v8 (Graphpad, San Diego, CA).

Cell Line

The patient derived SNUC cell line, MDA8788-6, was generously provided by MD Anderson. Generation of this cell line was previously described by Takahashi *et.al* (18). Cells were cultured in a humidified incubator at 37 °C with 5% (vol/vol) CO₂ in DMEM with 10% FBS, 1X Pen/Strep, 1X NEAA. Cells were genotyped to confirm the STR profile of the cell line (**Supplemental Table 2**) as previously described (21).

DNA Isolation

DNA was isolated from formalin fixed, paraffin-embedded (FFPE) samples following the manufacturer's protocol for AllPrep DNA/RNA FFPE kit (Qiagen, Hilden, Germany) as previously described (22, 23). Tumor and adjacent normal regions were identified on H&E stained slides and aligned to tissue paraffin blocks. An 18-gauge sterile needle was used to core 2–4 samples from each region. Deparaffinization was performed using the xylene/ethanol method with the only modification being that samples were digested using proteinase K at 56°C for 20–24 minutes. DNA isolation was then completed using the Allprep Isolation kit (Qiagen, Hilden, DE) following manufacturer protocol. Each sample was analyzed using a Nanodrop spectrophotometer for purity (260:230 and 260:280 ratios) and concentration was determined using 1uL of sample with the Qubit 2.0 Fluorometer and measured with a bioanalyzer as described (24). DNA extraction for MDA8788-6 cell line was performed using Wizard® Genomic DNA Purification Kit (Promega, Madison, WI).

DNA Sequencing

Genomic DNA from each tumor and adjacent normal specimen was submitted for sequencing to the University of Michigan's DNA sequencing core for exome sequencing using both the DNA TruSeq Exome Library Preparation kit (Illumina, Catalogue number FC-150-100x; SNUC2, SNUC5, SNUC8, SNUC10) and the Roche NimbleGen V3 capture kit (SNUC1). DNA from the MDA 8788-6 cell line was sequenced as described (25). Libraries were prepared according to the manufacturer's instructions. Libraries were then paired end sequenced to 125 nucleotides as part of pool with an average of 4 samples per lane on an Illumina HiSEQ4000 yielding an average depth of greater than 90x per sample.

Exome Variant Calling

Quality of the sequencing reads was assessed using FastQC v.0.11.5. Because the reads had adapter contamination as well as a high k-mer content at the start of the reads, trim galore v0.4.4 was used to remove adapters and trim reads. Reads were aligned to the hg19 reference genome using BWA v0.7.1. Mapping was followed by marking duplicates using PicardTools v1.79. Base quality score recalibration was done using GATK v3.6 and this was the last step in preparing the reads for variant calling. Samtools v1.2 was used to create pileup files for each tumor-normal pair. Varscan v2.4.1 was used to call variants from these mpileup files using the somatic mode of the variant caller. Goldex Helix Varseq v1.4.6 was used to annotate variants. All variants in the introns and intergenic regions were filtered out. Variants with more than 5 reads supporting the alternate allele in the tumor samples were considered as true positives.

Copy Number Analysis

Aberration Detection in Tumour Exome (ADTEEx) v.2.0 was used to make copy number estimation calls from the pre-processed tumor-normal BAM files which were also used for variant calling. A state from 0 to 4 was assigned by the software based on its estimated copy number. State 0 corresponds to a homozygous deletion, 1 corresponds to a heterozygous deletion. A normal copy number is denoted by state 2. States 3 and 4 represent a gain and amplification respectively.

Microsatellite Instability (MSI) Detection

MSIsensor was used to detect somatic MSI loci from the tumor-normal sample pairs as described previously (26). The software assigns a status to each sample pair based on an instability score calculated based on a threshold of more or less than 3.5% of called microsatellites having alterations. We present this score as well as the overall percentage of microsatellite alterations for each tumor-normal pair.

Sanger Sequencing

Excess DNA from above was used to validate mutation calls in novel genes. Primers were designed using MITprimer3 to amplify a small region surrounding the nominated single nucleotide variants (SNVs) as described in **Supplemental Table 3**. Polymerase chain reactions (PCRs) were optimized for each primer pair on cell line genomic DNA and then used to amplify the regions from tumor and adjacent normal DNA using Platinum Taq DNA High Fidelity polymerase (ThermoFisher, Waltham, MA). PCR products were then visualized on an eGel (Invitrogen, Waltham, MA) and purified using the a PCR purification kit (Qiagen, Valenica, CA) as described (27) and submitted for Sanger sequencing at the University of Michigan's DNA sequencing core. Results were visualized using the LaserGene software suite.

Cell line RNA Sequencing and Analysis

Total RNA from the MDA8788-6 sample underwent standard QC and was submitted for RNA sequencing to the University's DNA sequencing core as previously described (28, 29). Briefly, the Illumina Stranded RNAseq kit was used and libraries were sequenced on an Illumina HiSEQ4000 using 75nt paired end approach. Quality of the RNA sequencing reads was determined using FastQC v0.11.5 and we did not identify any quality issues. We then used a two-step alignment protocol of Star v2.5.3a to map the reads and genome index files were first generated using the reference human genome and annotated transcriptome files. In the second step, we then used the index files to guide read mapping. Samtools v1.9 and Picard v2.4.1 were used to retain only uniquely mapped reads and FPKM was computed using Cufflinks v2.2.1 with default parameters, with the exception of modifying "--max-bundle-frags" to 100,000,000. This modification was made to avoid raising of the HIDATA flag at loci that have more fragments than the pre-set threshold for every locus.

Fusion Gene Analysis

FusionCatcher (v1.00) is a software package designed to look for gene fusions, translocations and rearrangement events using paired end RNA-Seq data and was used to identify novel gene fusions in the MDA8788-6 cell line.

Linked Read Sequencing

High molecular weight DNA was isolated from the SNUC cell line by lysing 1.5 million cells overnight at 37° with lysis buffer (10 mM Tris-HCl, 400 mM NaCl, 2 mM EDTA), 10% SDS, and a proteinase K solution (1 mg/mL Proteinase K, 1% SDS, 2 mM EDTA). Following overnight lysis, DNA was salted out of the solution with 5 M NaCl for 1 hour at 4° and precipitated with ice cold ethanol for 5 hours at -20 °C. High molecular weight DNA was eluted in TE buffer; the quality and integrity of the DNA was assessed using the TapeStation Genomic DNA ScreenTape kit (Agilent). The DNA was submitted to the University's DNA

sequencing core for 10x based linked read library generation and sequencing on an Illumina NovaSeq6000 with 300nt paired end run. Samples were de-multiplexed and FastQ files with matched index files were generated using Long Ranger Version 2.2.2. Data was visualized using Loupe software package, Version 2.1.1 (2.4).

Fusion Gene Knockdown

All siRNA including ON-TARGETplus Non-targeting Control siRNA, ON-TARGETplus GAPDH Control siRNA, and a custom siRNA targeting the PGAP3-SRPK1 fusion site were purchased through Dharmacon (Lafayette, CO). Each siRNA was reconstituted at a concentration of 1 nmol/50 uL in 1X siRNA buffer (Dharmacon, Lafayette, CO). MDA8788-6 cells were plated at a concentration of 250,000 cell per well in 3 mL growth media. The following day all media was removed and cells were starved in 1 mL of serum DMEM for 3 hours. Each siRNA was prepared by adding 400uL of OPTI-MEM (Gibco, Waltham, Massachusetts) with 24uL of siRNA and left to equilibrate for 5 minutes. Separately, 24uL of oligofectamine (Invitrogen, Carlsbad, CA) was added to 96uL of OPTI-MEM. After five minutes the two mixtures are added together and allowed to equilibrate at room temperature for 20 minutes. Cell were then treated with 250uL of siRNA mixture containing buffer only, Non-targeting siRNA, PGAP3-SRPK1 fusion siRNA, or GAPDH siRNA. After 3 hours 2.5 mL of growth medium was added to each well. The following day cells were harvested in 700uL of QIAzol Lysis Reagent (Qiagen, Valencia, CA) and proceeded directly to RNA extraction or stored at minus 80 °C for future extraction. RNA extraction was performed using RNeasy Mini Kit (Qiagen, Valencia, CA) per manufacturer's instructions. RNA sequencing of the fusion knockdown was also performed as above. Briefly, extracted total RNA from MDA8788-6 NT siRNA and PGAP3-SRPK1 fusion siRNA were submitted to the University's DNA sequencing core and processed as above (Illumina Stranded RNAseq kit was used and libraries were sequenced on an Illumina NovaSEQ6000 using 300 cycle paired end approach).

Quantitative Polymerase Chain Reaction (qPCR)

Confirmation of successful siRNA knockdown and validation of RNAseq findings were performed with qPCR. Following RNA extraction, cDNA synthesis was performed using SuperScript™ III First-Strand Synthesis System (Invitrogen, Carlsbad, CA) and qPCR was performed using QuantiTect SYBR Green PCR Kit (Qiagen, Valencia, CA) and run on QuantStudio5 (Applied Biosystems, Foster City CA). Targets included *SRPK1*, *PGAP3-SPRK1* fusion, *GAPDH*, *HSDL2*, *CCND1*, *FOXO4*, Beta-Actin, *HRPT*, and *RPL-19*; primer sequences are listed in **Supplemental Table 4**. Analysis was performed using the $2^{\Delta\Delta-Ct}$ method (30).

Results

Survival Analysis

Forty-six patients were included in the survival analysis. The median age at diagnosis 53 years with a range from 19–87 years. Median follow up time was 28 months (range < 1 month – 23 years). Two patients (4.3%) were treated with surgery alone, 21 patients (44%) with surgery in addition to adjuvant

radiation, chemotherapy, or both (CRT) and 23 patients (49%) with CRT alone. Twenty-five patients (53%) had persistent or recurrent disease after treatment. The median time to recurrence was 2.8 months. Of these 64% were locoregional recurrence or persistence while 32% were distant failures. Five-year overall survival was 42% (95% CI 27–56%) and 5-year disease specific survival was 46% (95% CI 30–61%) as seen in Fig. 1a. There was no significant difference in 5-year DSS when stratifying by use of surgery +/- CRT and CRT alone (50% [95% CI 27–69%] vs. 49% [95% CI 25–68%], $p = 0.85$), Fig. 1b. Multivariate analysis was performed using backward Wald cox regression. Variables included in the model were T-stage, nodal disease, tobacco use and treatment type (surgery +/- CRT vs. CRT alone). Only tobacco use was found to be significantly associated with decreased survival (HR 5.1 [95% CI 1.7–15], $p = 0.004$).

Figure 1: Kaplan-Meier Survival Curves. A) Five-year disease specific survival (DSS) and overall survival (OS) for SNUC cohort. B) Five-year DSS stratifying by the use of surgery with or without chemotherapy and radiation (CRT) compared to CRT alone.

Exome Sequencing of SNUC Tumors

We were able to isolate high quality DNA that met our quality control standards for sequencing from 5 retrospective SNUC and matched normal samples that were advanced for full exome sequencing. Within our 5 sample cohort, we had 2 patients that died within two years of diagnosis and 3 patients that survived for more than 5 years after diagnosis (Table 1). Using these tumor and adjacent normal DNA samples, we sequenced exomes to an average depth of 256,352,493 yielding an average of 218,566,135 uniquely mapped reads in each sample, for an average coverage of > 100X per tumor (**Supplemental Table 5**).

Table 1
Patient Demographics

Sample ID	Tobacco Use (Current, Former, Never)	Site	TNM Classification	Initial Treatment	Died of Disease (Y/N)	Survival Months
SNUC1	Never	Nasal Cavity	T4bN0M0	Surgery, Adjuvant RT	N	144
SNUC2	Former	Maxillary Sinus	T4aN0M0	CRT	Y	< 1
SNUC5	Never	Nasal Cavity	T3N0M0	Surgery, Adjuvant CRT	N	125
SNUC8	Never	Ethmoid Sinus	T4bN0M0	Surgery, Adjuvant CRT	N	130
SNUC10	Never	Nasal Cavity	T4bN0M0	Surgery, Adjuvant CRT	Y	3

Using this data, we first generated copy number calls for broad genomic regions and assessed the global view of the copy number variation in Circos plots (Fig. 2a, **Supplemental Tables 6A-E**). These plots demonstrated that 3/5 tumors showed a much higher level of copy number variation as compared to the other samples with frequent high level amplifications in chromosomes 4, 17, 19 and 22. Genes with copy gain or amplifications in multiple tumors included: *CD300C* which plays a role in innate immunity and antigen presentation via MHC class I and is a negative regulator of CD4 and CD8 T cells (31), *JAK3*, *E2F4* and *GLI1*, which have canonical and non-cannonical roles in tumor cell proliferation respectively (32, 33), *MDM2* which targets p53 for degradation (34), and *MLL2* which may play a role in epithelial-mesenchymal transition (EMT) (35). We also identified copy gains in *ERBB2* in 3/5 samples (Fig. 2b) suggesting a potential role of *ERBB2* function in a subset of SNUCs, which has been previously noted (18). Moreover, we also identified copy number losses in *SMARC* family genes (*SMARCA1*, *SMARCA2*, *SMARCA5*, *SMARCB1*, *SMARCC1* and *SMARCE1*) in numerous samples consistent with previous reports of the recurrence of alterations to this gene family in SNUC (16). Focal amplifications were identified in *WEE1*, *FGFR3* and *MAPK15* while focal copy losses were identified in *FAT1* and *SMARCA2* (Fig. 2c).

Analysis of SNV data revealed an average of 23.6 (range of 5–63) high confidence non-synonymous somatic alterations per tumor (**Supplemental Fig. 1**). Oncoplots for recurrently altered as well as targetable or cancer-relevant genes are shown in Fig. 3a and an extended list of the mutations with functional annotations and pathogenicity scores is shown in (**Supplementary Tables 7–9**). From this analysis, we identified a mutation in *IDH2* (Arg172Gly), which is a gene previously associated with SNUC (15, 17), as well as new mutation disrupting *SMARCAL1* (Thr742Met; Fig. 3b,c) suggesting alternative pathways may be implicated in tumorigenesis in these samples. Further, we identified potentially targetable alterations including an *ALK* Gly872Ser (Fig. 3d). We also identified two different ABC transporter genes, *ABCA10* (Glu507Ter) and *ABCB7* (GLU342GLN), suggesting a general pathway for drug resistance that may explain why some tumors respond poorly to chemotherapy.

Finally, we characterized microsatellite instability in each of the tumors using the MANTIS software package, which assigns an instability score to each tumor/normal pair, and determine that all of the SNUC patient samples were microsatellite stable (**Supplemental Table 10**).

Exome and RNAseq Analysis of SNUC Cell Line (MDA8788-6)

Our colleagues recently derived the first SNUC cell line, MDA8788-6, and we performed exome sequencing on this cell line model with subsequent SNV annotation using our previously established informatics pipelines for head and neck cell lines without available matched normal DNA. This sequencing yielded 225,772,526 reads, of which 99.7% uniquely mapped to the genome yielding an average coverage of > 100X (**Supplemental Table 11**). As no matched normal was available, joint calling was completed with our UM-SCC cell lines as described (25, 28), and filtering to retain only heterozygous calls and remove any SNV previously reported in dbSNP, yielded 563 potential single nucleotide variants, of which 182 were categorized as missense, frameshift or stopgains (**Supplemental Table 12**). Similar filtering for insertion/deletion calls yielded 779 potential INDELS, of which 39 were categorized as frameshift INDELS (**Supplemental Table 13**). Among these alterations, we identified *KMT2B* S421P and *NOTCH1* T2483

missense mutations, which may have an important role in pathogenesis of this tumor model given the previously established role of these genes in tumorigenesis.

Next, we performed paired end RNA sequencing to discover potential gene fusions in the cell line using the FusionCatcher algorithm. This analysis called 107 potential gene fusions in the cell line to a variety of known oncogenes and previously described fusion genes from other cancers (**Supplemental Table 14**). Importantly, we were able to validate the presence of a novel fusion between *PGAP3* exon 2 and *SRPK1* exon 13 (Fig. 4a). The resulting fusion gene creates an in-frame fusion gene predicted to encode a 244 amino acid, with predicted pI of 5.87 and molecular weight of 27.6 kDa protein. The resulting fusion protein is predicted to retain the SRPK1 protein kinase domain, which normally regulates constitutive and alternative pre-mRNA splicing machinery through phosphorylation of SR proteins (36).

To then test the hypothesis that large scale structural rearrangements drove fusion formation, we performed 10X linked read genome sequencing on high molecular weight DNA isolated from the MDA8788-6 cell line. We obtained 848,842,686 total reads and averaged 36.9x coverage, with 96.1% of the DNA molecules of > 20 kb in length and 58% greater than 100 kb (**Supplemental Table 15**). Long ranger pipeline analysis phased 99.2% of SNPs resulting in a maximum phase block of 47.9 MB (Fig. 4b). The analysis identified 888 large structural variant calls in the genome. Contrary to our hypothesis, there were no structural rearrangements around *PGAP3* or *SRPK1* identified (Fig. 4c-e) suggesting that the fusion is formed at the RNA level, possibly by trans-splicing events. However, we did identify structural alterations involving *ZNF546* and *AXL* (Chr19:40,490,000 and Chr19:41,760,000, 1.27 MB duplication, quality score 1000) as well as *CREB3L2* and *BRAF* (Chr7:137,630,000 and Chr7:140,490,000, 2.86 MB duplication) suggesting a potential role for these oncogenes in pathogenesis of this SNUC model (**Supplemental Fig. 2**).

Finally, to test for potential functional roles of the *PGAP3-SRPK1* fusion, we performed siRNA-mediated knockdown of the fusion transcript and submitted RNA for complete transcriptome sequencing. This data demonstrated that knockdown of *PGAP3-SRPK1* fusion results in significantly decreased *HSDL2*, *NAGK*, and *CCND1* RNA expression and a modest increase in *LINC01006* and *FOXO4* suggesting *PGAP3-SRPK1* may play a role in regulation of these genes (Fig. 5a). In fact, we identified 122 genes that were > 2-fold upregulated and 112 genes > 2-fold downregulated in the knockdown relative to control. Further, gene set enrichment analysis of the differentially expressed rank list found enrichment of KOBAYASHI_EGFR_SIGNALING_24HR_DN, HALLMARK_E2F_TARGETS, and HALLMARK_MYC_TARGETS_V1 pathways (Fig. 5b, **Supplemental Table 16**). Collectively, this RNA sequencing data analysis suggests that the gene fusion has a pivotal role in the hallmark signaling pathways.

Discussion

Survival analysis from our cohort is congruent with previous reports of low survival rates (1, 4–9) and shows little differences in survival when stratifying by treatment modality. In this study, we confirm the

presence of previously noted alterations in *ERBB2*, *IDH2*, and *SMARC* family members from initial targeted sequencing studies. Notably, we did not find a high frequency of *SMARCB1* mutations that have been previously identified, and only noted one *IDH2* mutation in our samples, suggesting genetic diversity for SNUC etiology. However, our data, identified not only copy number alterations in *SMARCB1*, but also to additional SMARC family members, and suggest that deregulation of the *SWI/SNF* nucleosome remodeling complex (consisting of known tumor suppressors *SMARCB1*, *SMARCA4*, *PBRM1*, *ARID1B*, and *ARID2*), through one of its many components, is a critical step in disease progression in SNUCs (37). *SMARCB1* has been implicated in numerous other solid cancers as a tumor suppressor gene, including sarcomas, carcinomas and rhabdoid tumors of varying sites (38). It appears to have tumor suppressor functions in inhibiting cell cycle and proliferation via the p16-Rb-E2F and Wnt/Beta-catenin pathways, among others (38). *SMARCA2* function and expression may also play a critical role in response to specific targeted therapies (particularly with EZH2 inhibition) in tumors with *SWI/SNF* dysregulation (39), suggesting a potential role for EZH2 inhibitors in SNUCs.

Previous isolated reports of SNUCs have identified overexpression or amplification of growth factor receptors (18, 19, 40) and in this study, we have identified genetic alterations in *ERBB2* and *FGFR* family growth factor receptor genes as well as *ALK*, suggesting potential targetable option in SNUCs. In a previous study of a SNUC cell line, high *ERBB2* expression was identified with a notable response to ERBB2 inhibition (18). *FGFR3* alterations have been implicated in head and neck cancer and *in vitro* and *in vivo* studies suggest a promising role for FGF inhibition in head and neck tumors (41–43). Given the diverse, but potentially actionable set of alterations that our data defined, these results suggest a role for in depth molecular analysis of this rare disease in order to gain insight into molecular alterations that may drive discovery of future therapeutics, and potentially guide individual patient treatment options.

Finally, this study identifies a novel fusion of *PGAP3-SRPK*. *SRPK1* has been previously characterized to drive cell proliferation, migration, and invasion in colorectal and gastric cancers (44–47) suggesting that the fusion protein may have oncogenic consequences in the SNUC cell line. CNV analysis additionally revealed copy gain in one tumor in the *SRPK1* gene. Knockdown of the PGAP3-SRPK1 fusion gene resulted in changes in expression of *CCND1*, *FOXO4* and most significantly a decrease in *HSDL2* and *NAGK* suggesting a functional role for this novel fusion gene. Unfortunately, insufficient RNA prevented evaluation of SNUC tissues for presence of the fusion.

Conclusion

Given the rarity of this tumor, it will be challenging to characterize a large cohort of patients. Nevertheless, we believe this initial analysis of five SNUCs represents a valuable preliminary guideline of the mutational landscape of SNUCs and identifies multiple recurring mutations and pathway alterations. These may be of particular interest both as prognostic biomarkers in larger cohort studies, and as potentially targetable therapeutic options. Consequently, the alterations identified here represent promising targets for future SNUC studies and support a potential pathogenic role in other cancers. Due to both the infrequency and

highly aggressive nature SNUC, our hope is this study will serve as primer to advance therapeutic concepts developed for other malignancies against these pathways into future SNUC therapies.

Declarations

Ethics approval and consent to participate – This is a retrospective data review study without direct involvement of human subjects and as such informed consent was waived by the University of Michigan Institutional Review Board (ID#: HUM00080561 and HUM00085888) to obtain surplus tissues from Department of Pathology archives for study. All methods were carried out in accordance with relevant guidelines and regulations from the University of Michigan ethics committee. All experimental protocols were approved by University of Michigan ethics committee.

Consent for Publication – Not applicable.

Availability of Data - All data generated or analyzed during this study are included in this published article and its supplementary information files as well as deposited online (SRA PRJNA668767).

Competing Interests - The authors declare that they have no competing interests

Funding - M.E.H received funding from NIH grant T32 DC005356. J.C.B. received funding from NIH Grants U01-DE025184, P30-CA046592 and R01-CA194536. M.L. was funded by F31-CA206341-01A1. Funding sources had no role in in the design of the study and collection, analysis, and interpretation of data or in writing the manuscript.

Author Contributions – MEH performed data collection, survival analysis, sequencing analysis, data visualization, siRNA knockout experiments, drafted manuscript, edited manuscript, and approved final version, ACB analyzed sequencing data, drafted manuscript, edited manuscript, and approved final version, ADB performed bioinformatics analysis and contributed to data visualization, edited manuscript, and approved final version, JZ performed formal statistical analysis, edited manuscript, and approved final version, AK performed bioinformatics analysis and contributed to data visualization, edited manuscript, and approved final version, SKF performed DNA and RNA extraction, prepped DNA and RNA for sequencing submission, contributed to data visualization, edited manuscript, and approved final version, BMJ performed DNA and RNA extraction, prepped DNA and RNA for sequencing submission, contributed to data visualization, edited manuscript, and approved final version, MLL contributed to sequencing analysis, data visualization, edited manuscript, and approved final version, LP performed DNA preparation and submission for linked read sequencing, contributed to data analysis and data visualization, edited manuscript, and approved final version, JH performed formal statistical analysis, edited manuscript, and approved final version, LM performed data collection, edited manuscript, and approved final version, ELM provided supervision and contributed to project design, performed data collection, edited manuscript, and approved final version, JCB is responsible for project design, supervision, data analysis, data visualization, editing manuscript, and approval of final version.

References

1. Chambers KJ, Lehmann AE, Remenschneider A, Dedmon M, Meier J, Gray ST, Lin DT. Incidence and survival patterns of sinonasal undifferentiated carcinoma in the United States. *Journal of neurological surgery Part B, Skull base*. 2015;76(2):94-100. doi: 10.1055/s-0034-1390016. PubMed PMID: 25844294; PMCID: 4375051.
2. Ashraf MJ, Azarpira N, Khademi B, Abedi E, Hakimzadeh A, Valibeigi B. Extramedullary plasmacytoma of the nasal cavity report of three cases with review of the literature. *Iranian Red Crescent medical journal*. 2013;15(4):363-6. Epub 2013/10/02. doi: 10.5812/ircmj.2209. PubMed PMID: 24083014; PMCID: PMC3785915.
3. Bridge JA, Bowen JM, Smith RB. The small round blue cell tumors of the sinonasal area. *Head and neck pathology*. 2010;4(1):84-93. Epub 2010/03/20. doi: 10.1007/s12105-009-0158-6. PubMed PMID: 20237994; PMCID: PMC2825526.
4. Gelbard A, Hale KS, Takahashi Y, Davies M, Kupferman ME, El-Naggar AK, Myers JN, Hanna EY. Molecular profiling of sinonasal undifferentiated carcinoma. *Head & neck*. 2014;36(1):15-21. doi: 10.1002/hed.23267. PubMed PMID: 23633104; PMCID: 3874284.
5. Lin EM, Sparano A, Spalding A, Eisbruch A, Worden FP, Heth J, Sullivan SE, Thompson BG, Marentette LJ. Sinonasal undifferentiated carcinoma: a 13-year experience at a single institution. *Skull base : official journal of North American Skull Base Society [et al]*. 2010;20(2):61-7. Epub 2010/09/03. doi: 10.1055/s-0029-1236165. PubMed PMID: 20808529; PMCID: PMC2853070.
6. Mendenhall WM, Mendenhall CM, Riggs CE, Jr., Villaret DB, Mendenhall NP. Sinonasal undifferentiated carcinoma. *American journal of clinical oncology*. 2006;29(1):27-31. Epub 2006/02/08. doi: 10.1097/01.coc.0000189691.04140.02. PubMed PMID: 16462499.
7. Miyamoto RC, Gleich LL, Biddinger PW, Gluckman JL. Esthesioneuroblastoma and sinonasal undifferentiated carcinoma: impact of histological grading and clinical staging on survival and prognosis. *The Laryngoscope*. 2000;110(8):1262-5. Epub 2000/08/15. doi: 10.1097/00005537-200008000-00007. PubMed PMID: 10942123.
8. Musy PY, Reibel JF, Levine PA. Sinonasal undifferentiated carcinoma: the search for a better outcome. *The Laryngoscope*. 2002;112(8 Pt 1):1450-5. Epub 2002/08/13. doi: 10.1097/00005537-200208000-00023. PubMed PMID: 12172261.
9. Rischin D, Porceddu S, Peters L, Martin J, Corry J, Weih L. Promising results with chemoradiation in patients with sinonasal undifferentiated carcinoma. *Head & neck*. 2004;26(5):435-41. Epub 2004/05/04. doi: 10.1002/hed.10396. PubMed PMID: 15122660.
10. Gray ST, Herr MW, Sethi RK, Diercks G, Lee L, Curry W, Chan A, Clark J, Holbrook EH, Rocco J, Sadow PM, Lin DT. Treatment outcomes and prognostic factors, including human papillomavirus, for sinonasal undifferentiated carcinoma: a retrospective review. *Head & neck*. 2015;37(3):366-74. doi: 10.1002/hed.23606. PubMed PMID: 24421248.

11. Morand GB, Anderegg N, Vital D, Ikenberg K, Huber GF, Soyka MB, Egger M, Holzmann D. Outcome by treatment modality in sinonasal undifferentiated carcinoma (SNUC): A case-series, systematic review and meta-analysis. *Oral oncology*. 2017;75:28-34. doi: 10.1016/j.oraloncology.2017.10.008. PubMed PMID: 29224819.
12. Gamez ME, Lal D, Halyard MY, Wong WW, Vargas C, Ma D, Ko SJ, Foote RL, Patel SH. Outcomes and patterns of failure for sinonasal undifferentiated carcinoma (SNUC): The Mayo Clinic Experience. *Head & neck*. 2017;39(9):1819-24. doi: 10.1002/hed.24834. PubMed PMID: 28561906.
13. Kuo P, Manes RP, Schwam ZG, Judson BL. Survival Outcomes for Combined Modality Therapy for Sinonasal Undifferentiated Carcinoma. *Otolaryngology–head and neck surgery : official journal of American Academy of Otolaryngology-Head and Neck Surgery*. 2017;156(1):132-6. doi: 10.1177/0194599816670146. PubMed PMID: 27703092.
14. Amit M, Abdelmeguid AS, Watcherporn T, Takahashi H, Tam S, Bell D, Ferrarotto R, Glisson B, Kupferman ME, Roberts DB, Su SY, Raza SM, DeMonte F, Hanna EY. Induction Chemotherapy Response as a Guide for Treatment Optimization in Sinonasal Undifferentiated Carcinoma. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2019;37(6):504-12. Epub 2019/01/08. doi: 10.1200/jco.18.00353. PubMed PMID: 30615549; PMCID: PMC6380524.
15. Dogan S, Chute DJ, Xu B, Ptashkin RN, Chandramohan R, Casanova-Murphy J, Nafa K, Bishop JA, Chiosea SI, Stelow EB, Ganly I, Pfister DG, Katabi N, Ghossein RA, Berger MF. Frequent IDH2 R172 mutations in undifferentiated and poorly-differentiated sinonasal carcinomas. *The Journal of pathology*. 2017;242(4):400-8. doi: 10.1002/path.4915. PubMed PMID: 28493366; PMCID: 5639875.
16. Agaimy A, Hartmann A, Antonescu CR, Chiosea SI, El-Mofty SK, Geddert H, Iro H, Lewis JS, Jr., Markl B, Mills SE, Riener MO, Robertson T, Sandison A, Semrau S, Simpson RH, Stelow E, Westra WH, Bishop JA. SMARCB1 (INI-1)-deficient Sinonasal Carcinoma: A Series of 39 Cases Expanding the Morphologic and Clinicopathologic Spectrum of a Recently Described Entity. *The American journal of surgical pathology*. 2017;41(4):458-71. doi: 10.1097/PAS.0000000000000797. PubMed PMID: 28291122; PMCID: 5354087.
17. Jo VY, Chau NG, Hornick JL, Krane JF, Sholl LM. Recurrent IDH2 R172X mutations in sinonasal undifferentiated carcinoma. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*. 2017;30(5):650-9. doi: 10.1038/modpathol.2016.239. PubMed PMID: 28084339.
18. Takahashi Y, Lee J, Pickering C, Bell D, Jiffar TW, Myers JN, Hanna EY, Kupferman ME. Human epidermal growth factor receptor 2/neu as a novel therapeutic target in sinonasal undifferentiated carcinoma. *Head & neck*. 2016;38 Suppl 1:E1926-34. doi: 10.1002/hed.24350. PubMed PMID: 26752332.
19. Schrock A, Goke F, Wagner P, Bode M, Franzen A, Huss S, Agaimy A, Ihrler S, Kirsten R, Kristiansen G, Bootz F, Lengerke C, Perner S. Fibroblast growth factor receptor-1 as a potential therapeutic target in sinonasal cancer. *Head & neck*. 2014;36(9):1253-7. doi: 10.1002/hed.23443. PubMed PMID: 23913758.

20. Shuman AG, Gornick MC, Brummel C, Kent M, Spector-Bagdady K, Biddle E, Bradford CR, Brenner JC. Patient and Provider Perspectives Regarding Enrollment in Head and Neck Cancer Research. *Otolaryngol Head Neck Surg.* 2020;162(1):73-8. Epub 2019/12/11. doi: 10.1177/0194599819889976. PubMed PMID: 31818193; PMCID: PMC6946860.
21. Brenner JC, Graham MP, Kumar B, Saunders LM, Kupfer R, Lyons RH, Bradford CR, Carey TE. Genotyping of 73 UM-SCC head and neck squamous cell carcinoma cell lines. *Head & neck.* 2010;32(4):417-26. Epub 2009/09/18. doi: 10.1002/hed.21198. PubMed PMID: 19760794; PMCID: PMC3292176.
22. Birkeland AC, Burgin SJ, Yanik M, Scott MV, Bradford CR, McHugh JB, McLean SA, Sullivan SE, Nor JE, McKean EL, Brenner JC. Pathogenetic Analysis of Sinonasal Teratocarcinosarcomas Reveal Actionable beta-catenin Overexpression and a beta-catenin Mutation. *J Neurol Surg B Skull Base.* 2017;78(4):346-52. Epub 2017/07/21. doi: 10.1055/s-0037-1601320. PubMed PMID: 28725522; PMCID: PMC5515660.
23. Smith J, Kulkarni A, Birkeland AC, McHugh JB, Brenner JC. Whole-Exome Sequencing of Sinonasal Small Cell Carcinoma Arising within a Papillary Schneiderian Carcinoma In Situ. *Otolaryngol Head Neck Surg.* 2018;159(5):859-65. doi: 10.1177/0194599818774004. PubMed PMID: 29734873; PMCID: PMC6212311.
24. Birkeland AC, Foltin SK, Michmerhuizen NL, Hoesli RC, Rosko AJ, Byrd S, Yanik M, Nor JE, Bradford CR, Prince ME, Carey TE, McHugh JB, Spector ME, Brenner JC. Correlation of Crtc1/3-Maml2 fusion status, grade and survival in mucoepidermoid carcinoma. *Oral Oncol.* 2017;68:5-8. doi: 10.1016/j.oraloncology.2017.02.025. PubMed PMID: 28438292; PMCID: 5433350.
25. Mann JE, Kulkarni A, Birkeland AC, Kafelghazal J, Eisenberg J, Jewell BM, Ludwig ML, Spector ME, Jiang H, Carey TE, Brenner JC. The molecular landscape of the University of Michigan laryngeal squamous cell carcinoma cell line panel. *Head & neck.* 2019. doi: 10.1002/hed.25803. PubMed PMID: 31090975.
26. Niu B, Ye K, Zhang Q, Lu C, Xie M, McLellan MD, Wendl MC, Ding L. MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. *Bioinformatics (Oxford, England).* 2014;30(7):1015-6. Epub 2013/12/29. doi: 10.1093/bioinformatics/btt755. PubMed PMID: 24371154; PMCID: PMC3967115.
27. Michmerhuizen NL, Leonard E, Matovina C, Harris M, Herbst G, Kulkarni A, Zhai J, Jiang H, Carey TE, Brenner JC. Rationale for Using Irreversible Epidermal Growth Factor Receptor Inhibitors in Combination with Phosphatidylinositol 3-Kinase Inhibitors for Advanced Head and Neck Squamous Cell Carcinoma. *Mol Pharmacol.* 2019;95(5):528-36. Epub 2019/03/13. doi: 10.1124/mol.118.115162. PubMed PMID: 30858165; PMCID: PMC6442321.
28. Ludwig ML, Kulkarni A, Birkeland AC, Michmerhuizen NL, Foltin SK, Mann JE, Hoesli RC, Devenport SN, Jewell BM, Shuman AG, Spector ME, Carey TE, Jiang H, Brenner JC. The genomic landscape of UM-SCC oral cavity squamous cell carcinoma cell lines. *Oral Oncol.* 2018;87:144-51. doi: 10.1016/j.oraloncology.2018.10.031. PubMed PMID: 30527230; PMCID: PMC6349383.

29. Mann JE, Kulkarni A, Birkeland AC, Kafelghazal J, Eisenberg J, Jewell BM, Ludwig ML, Spector ME, Jiang H, Carey TE, Brenner JC. The molecular landscape of the University of Michigan laryngeal squamous cell carcinoma cell line panel. *Head & neck*. 2019;41(9):3114-24. Epub 2019/05/16. doi: 10.1002/hed.25803. PubMed PMID: 31090975; PMCID: PMC6692232.
30. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods (San Diego, Calif)*. 2001;25(4):402-8. Epub 2002/02/16. doi: 10.1006/meth.2001.1262. PubMed PMID: 11846609.
31. Cui C, Su M, Lin Y, Lai L. A CD300c-Fc Fusion Protein Inhibits T Cell Immunity. *Frontiers in immunology*. 2018;9:2657. Epub 2018/12/01. doi: 10.3389/fimmu.2018.02657. PubMed PMID: 30498497; PMCID: PMC6249344.
32. Hsu J, Sage J. Novel functions for the transcription factor E2F4 in development and disease. *Cell cycle (Georgetown, Tex)*. 2016;15(23):3183-90. Epub 2016/10/19. doi: 10.1080/15384101.2016.1234551. PubMed PMID: 27753528; PMCID: PMC5176148.
33. Mastrangelo E, Milani M. Role and inhibition of GLI1 protein in cancer. *Lung Cancer (Auckland, NZ)*. 2018;9:35-43. Epub 2018/04/10. doi: 10.2147/lctt.S124483. PubMed PMID: 29628779; PMCID: PMC5877502.
34. Shangary S, Wang S. Targeting the MDM2-p53 interaction for cancer therapy. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2008;14(17):5318-24. Epub 2008/09/04. doi: 10.1158/1078-0432.Ccr-07-5136. PubMed PMID: 18765522; PMCID: PMC2676446.
35. Abudurehman A, Ainiwaer J, Hou Z, Niyaz M, Turghun A, Hasim A, Zhang H, Lu X, Sheyhidin I. High MLL2 expression predicts poor prognosis and promotes tumor progression by inducing EMT in esophageal squamous cell carcinoma. *Journal of cancer research and clinical oncology*. 2018;144(6):1025-35. Epub 2018/03/14. doi: 10.1007/s00432-018-2625-5. PubMed PMID: 29532228; PMCID: PMC5948284.
36. Bullock N, Oltean S. The many faces of SRPK1. *The Journal of pathology*. 2017;241(4):437-40. Epub 2016/11/20. doi: 10.1002/path.4846. PubMed PMID: 27859253; PMCID: PMC5324686.
37. Alver BH, Kim KH, Lu P, Wang X, Manchester HE, Wang W, Haswell JR, Park PJ, Roberts CW. The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers. *Nature communications*. 2017;8:14648. doi: 10.1038/ncomms14648. PubMed PMID: 28262751; PMCID: 5343482.
38. Kohashi K, Oda Y. Oncogenic roles of SMARCB1/INI1 and its deficient tumors. *Cancer science*. 2017;108(4):547-52. doi: 10.1111/cas.13173. PubMed PMID: 28109176; PMCID: 5406539.
39. Januario T, Ye X, Bainer R, Alicke B, Smith T, Haley B, Modrusan Z, Gould S, Yauch RL. PRC2-mediated repression of SMARCA2 predicts EZH2 inhibitor activity in SWI/SNF mutant tumors. *Proceedings of the National Academy of Sciences of the United States of America*. 2017;114(46):12249-54. doi: 10.1073/pnas.1703966114. PubMed PMID: 29087303; PMCID: 5699030.

40. Chernock RD, Perry A, Pfeifer JD, Holden JA, Lewis JS, Jr. Receptor tyrosine kinases in sinonasal undifferentiated carcinomas—evaluation for EGFR, c-KIT, and HER2/neu expression. *Head & neck*. 2009;31(7):919-27. doi: 10.1002/hed.21061. PubMed PMID: 19283847.
41. Ban MJ, Byeon HK, Yang YJ, An S, Kim JW, Kim JH, Kim DH, Yang J, Kee H, Koh YW. Fibroblast growth factor receptor 3-mediated reactivation of ERK signaling promotes head and neck squamous cancer cell insensitivity to MEK inhibition. *Cancer science*. 2018;109(12):3816-25. Epub 2018/10/22. doi: 10.1111/cas.13839. PubMed PMID: 30343534; PMCID: PMC6272115.
42. McDermott SC, Rodriguez-Ramirez C, McDermott SP, Wicha MS, Nor JE. FGFR signaling regulates resistance of head and neck cancer stem cells to cisplatin. *Oncotarget*. 2018;9(38):25148-65. Epub 2018/06/05. doi: 10.18632/oncotarget.25358. PubMed PMID: 29861860; PMCID: PMC5982758.
43. Sweeny L, Liu Z, Lancaster W, Hart J, Hartman YE, Rosenthal EL. Inhibition of fibroblasts reduced head and neck cancer growth by targeting fibroblast growth factor receptor. *The Laryngoscope*. 2012;122(7):1539-44. Epub 2012/03/31. doi: 10.1002/lary.23266. PubMed PMID: 22460537; PMCID: PMC3447628.
44. Yao Y, Li Q, Wang H. MiR-216b suppresses colorectal cancer proliferation, migration, and invasion by targeting SRPK1. *OncoTargets and therapy*. 2018;11:1671-81. doi: 10.2147/OTT.S161835. PubMed PMID: 29615842; PMCID: 5870636.
45. Li Q, Wang G, Wang H. MiR-126 functions as a tumor suppressor by targeting SRPK1 in human gastric cancer. *Oncology research*. 2018. doi: 10.3727/096504018X15180508535835. PubMed PMID: 29510776.
46. Wang H, Wang C, Tian W, Yao Y. The crucial role of SRPK1 in IGF-1-induced EMT of human gastric cancer. *Oncotarget*. 2017;8(42):72157-66. doi: 10.18632/oncotarget.20048. PubMed PMID: 29069776; PMCID: 5641119.
47. Xu X, Wei Y, Wang S, Luo M, Zeng H. Serine-arginine protein kinase 1 (SRPK1) is elevated in gastric cancer and plays oncogenic functions. *Oncotarget*. 2017;8(37):61944-57. doi: 10.18632/oncotarget.18734. PubMed PMID: 28977917; PMCID: 5617477.

Figures

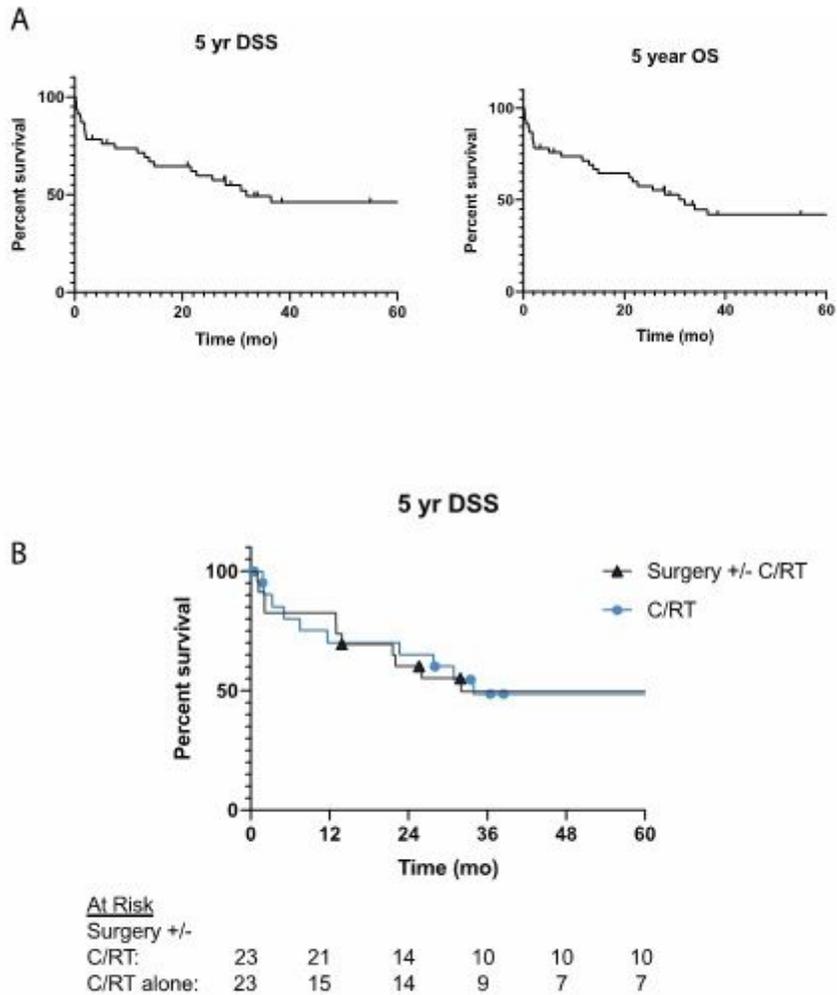


Figure 1

Kaplan-Meier Survival Curves. A) Five-year disease specific survival (DSS) and overall survival (OS) for SNUC cohort. B) Five-year DSS stratifying by the use of surgery with or without chemotherapy and radiation (CRT) compared to CRT alone.

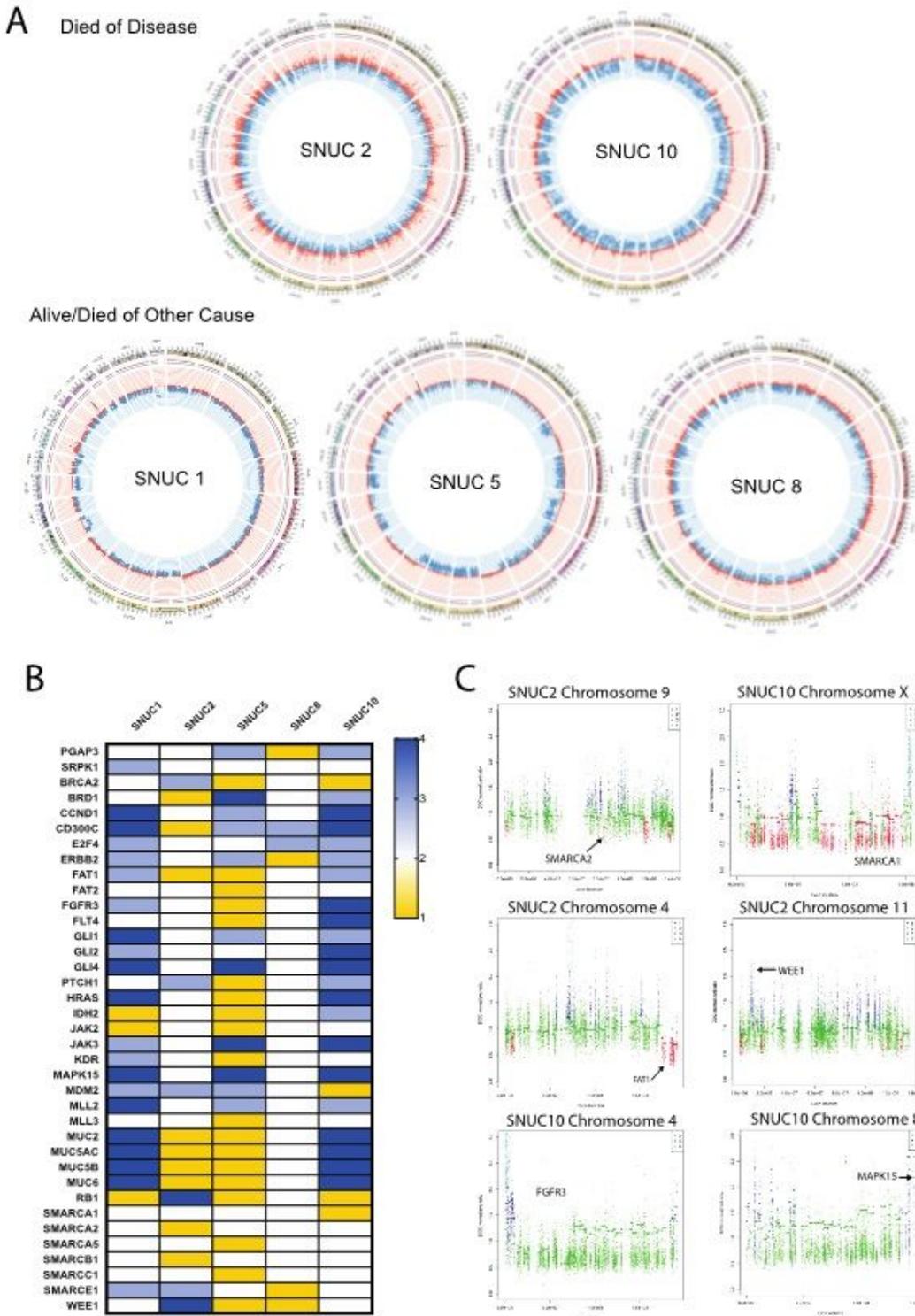


Figure 2

Copy Number Variation (CNV) Analysis. A) Circos Plots depicting copy number variation for each sample. For patients who dies of disease, 2/2 had increased copy number variations, compared to only 1/3 in the group of patients who did not die of disease. B) Heatmap depicting copy number for key genes; 1- copy loss, 2 – copy neutral, 3- copy gain, and 4 – amplification. C) Manhattan plots for highlighted genes with focal copy alterations.

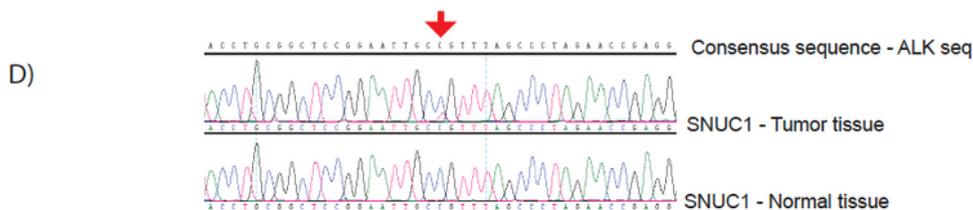
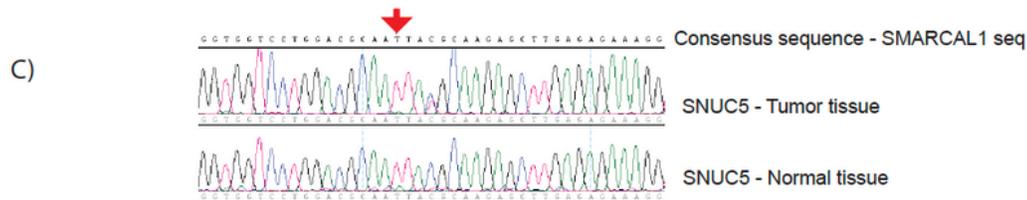
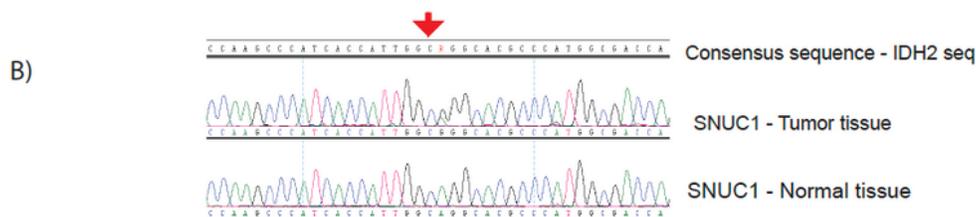
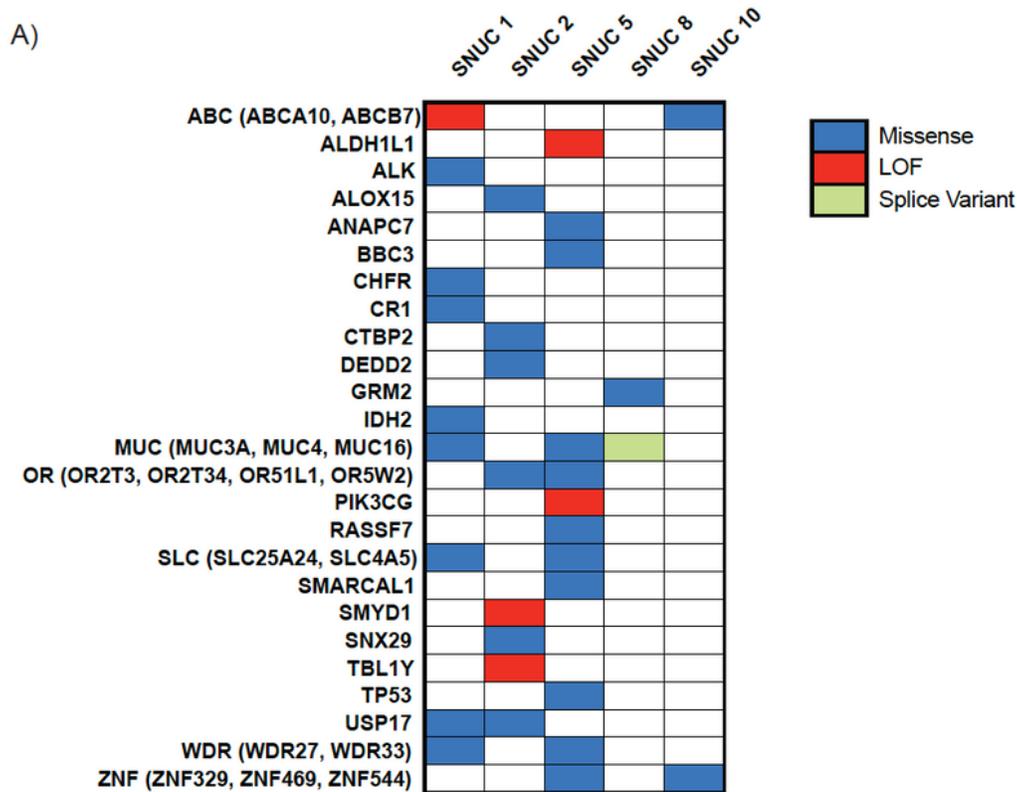


Figure 3

A) OncoPrint depicting high confidence non-synonymous somatic alterations in key genes in five SNUC samples. Sanger Sequencing Validation of Single Nucleotide Variations. B) IDH2 Arg172Gly in SNUC1. C) SMARCAL1 Thr742Met in SNUC5. D) ALK Gly872Ser.

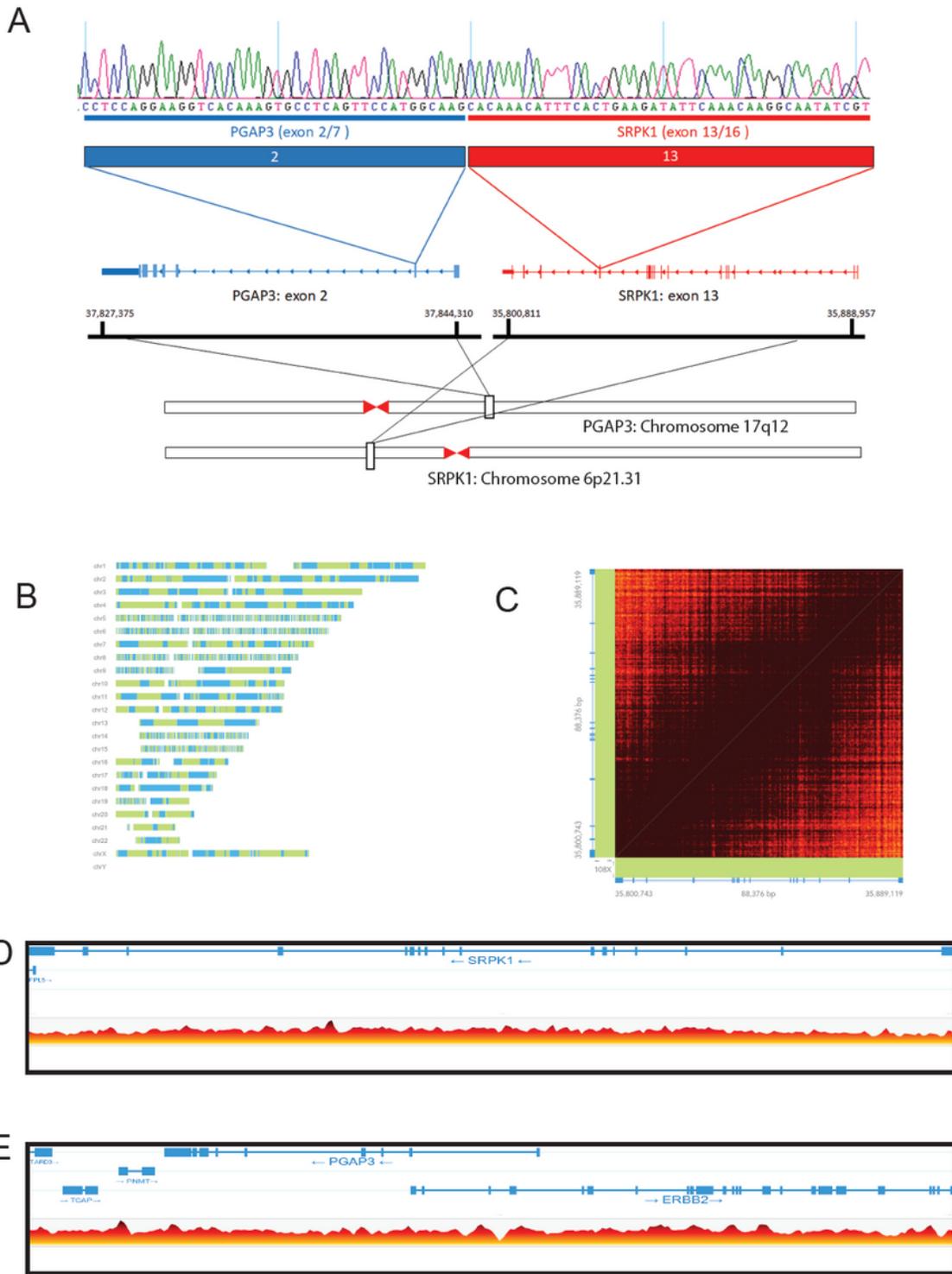
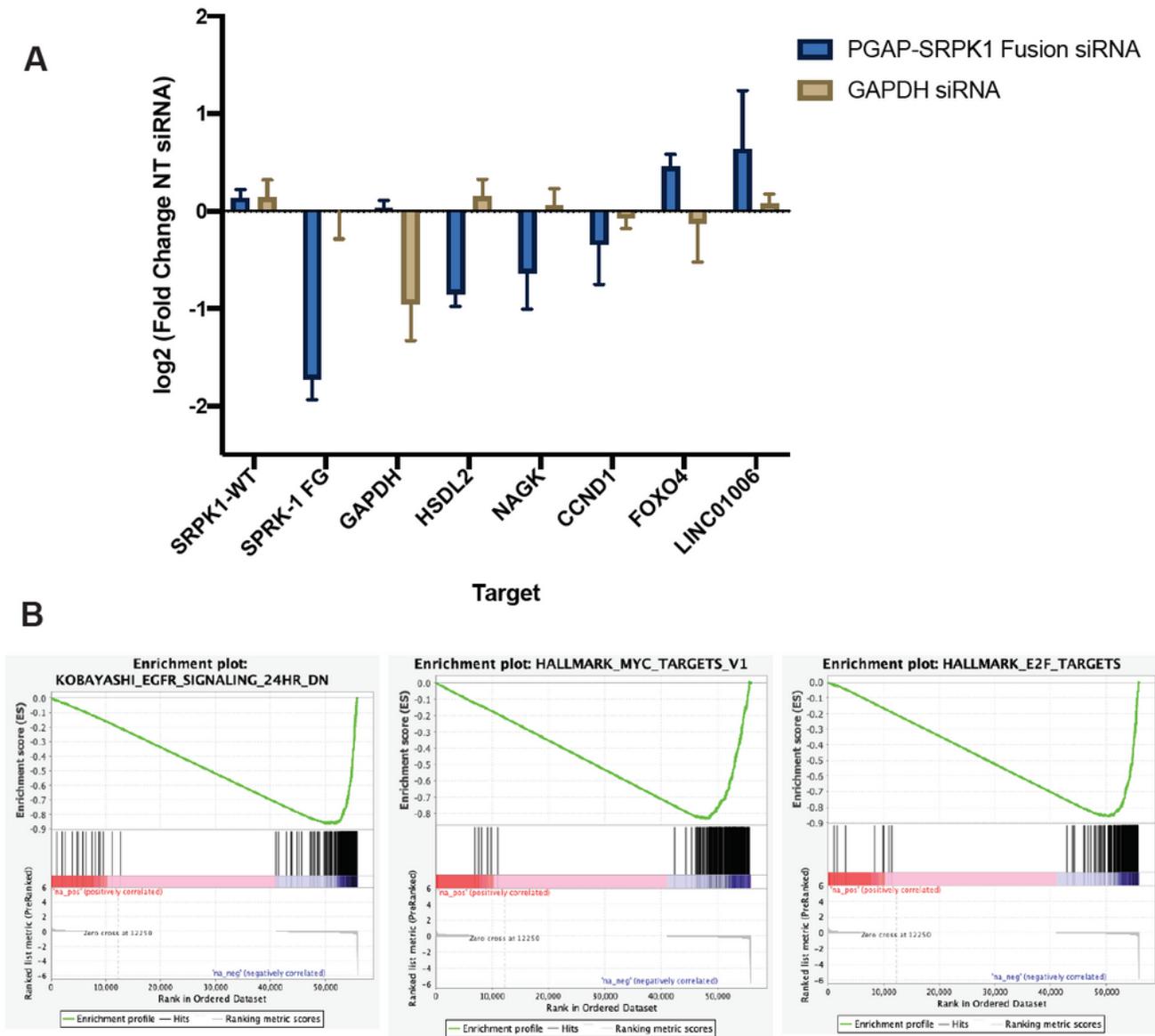


Figure 4

PGAP3-SRPK1 fusion gene and linked read genome sequencing. A) Schematic representation of the novel PGAP4-SRPK1 fusion gene. Exon 2 of PGAP3 is fused to exon 13 of SPRK1. B) Phase map summarizing linked read sequencing data as reported by the LongRanger and Loupe Analysis pipeline. C) Linked read matrix map showing high density coverage of linked reads over the SRPK1 locus. Dark brown indicates >30x coverage. D,E) Linear view of linked read data highlighting structural variations (indicated

with colored lines, with horizontal bars indicating the direction of each paired read), for SRPK1 (D) and PGAP3 (E)



Number	Gene Set	Nominal p-value	FDR q-value
1	KOBAYASHI_EGFR_SIGNALING_24HR_DN	<0.001	<0.001
3	HALLMARK_E2F_TARGETS	<0.001	<0.001
14	HALLMARK_MYC_TARGETS_V1	<0.001	<0.001

Figure 5

RNAseq and gene set enrichment analysis (GSEA) results for PGAP3-SRPK1 fusion knockdown. A) qPCR of MD8788-6 treated with PGAP3-SRPK1 siRNA compared to NT siRNA. Results are quantified as log₂ fold change of NT siRNA samples. Results demonstrate knockdown of PGAP3/SRPK1 mRNA

without changes in WT SRPK1 expression and confirm RNAseq results showing decreases in HSDL2 and CCDN1 and a small increase in FOXO4 with PGAP3-SRPK1 knockdown. B) GSEA results demonstrating a negative correlation with KOBAYASHI_EGFR_SIGNALING_24HR_DN, HALLMARK_E2F_TARGETS, and HALLMARK_MYC_TARGETS_V1 pathways in samples with PGAP3-SRPK1 knockdown.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementalFigure1.docx](#)
- [SupplementalFigure2.pdf](#)
- [SupplementalTable1demographicstable.docx](#)
- [SupplementalTable2SnucSTRProfile.xlsx](#)
- [Supplementaltable3.docx](#)
- [SupplementalTable4primers.docx](#)
- [supplementaltable5WESqc.xlsx](#)
- [SupplementalTable6ACNVdata.xlsx](#)
- [SupplementalTable6BCNVdata.xlsx](#)
- [SupplementalTable6CCNVdata.xlsx](#)
- [SupplementalTable6DCNVdata.xlsx](#)
- [SupplementalTable6ECNVdata.xlsx](#)
- [SupplementalTable7SNV.xlsx](#)
- [SupplementalTable8Indels.xlsx](#)
- [SupplementalTable9VEST.xlsx](#)
- [SupplementalTable10MSI.xlsx](#)
- [SupplementalTable11CelllineExomeQC.xlsx](#)
- [SupplementalTable12SNUCSNVsannotatedexomefinal.xlsx](#)
- [Supplementaltable13SNUCINDELSannotatedexomefinal.xlsx](#)
- [SupplementalTable14SNUCCellLineGeneFusions.xlsx](#)
- [SupplementalTable15LongRangerData.xlsx](#)
- [Supplementaltable16SNUCknockdownGSEAsummary.xlsx](#)