nature medicine

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41591-019-0749-z

In the format provided by the authors and unedited.

The Angiosarcoma Project: enabling genomic and clinical discoveries in a rare cancer through patient-partnered research

Corrie A. Painter[1,2,15], Esha Jain[1,2,3,15], Brett N. Tomson[1,2,15], Michael Dunphy[1,2], Rachel E. Stoddard[1,2], Beena S. Thomas[1,2], Alyssa L. Damon[1,2], Shahrayz Shah[1,2], Dewey Kim[1,2,3], Jorge Gómez Tejeda Zañudo[2,3], Jason L. Hornick[4], Yen-Lin Chen[5], Priscilla Merriam[3,6], Chandrajit P. Raut[6,7], George D. Demetri[3,6,8], Brian A. Van Tine[9], Eric S. Lander[1,2,10,11], Todd R. Golub[1,2,12] and Nikhil Wagle[1,2,3,13,14]*

[1]Count Me In, Cambridge, MA, USA. [2]The Broad Institute of MIT and Harvard, Cambridge, MA, USA. [3]Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA. [4]Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [5]Department of Radiation Oncology, Massachusetts General Hospital, Boston, MA, USA. [6]Center for Sarcoma and Bone Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA. [7]Department of Surgery, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [8]Ludwig Center at Harvard, Harvard Medical School, Boston, MA, USA. [9]Division of Oncology, Washington University School of Medicine in St. Louis, St. Louis, MO, USA. [10]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA. [11]Department of Systems Biology, Harvard Medical School, Boston, MA, USA. [12]Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA. [13]Center for Cancer Genomics, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA. [14]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [15]These authors contributed equally: Corrie A. Painter, Esha Jain, Brett N. Tomson.
*e-mail: nikhil_wagle@dfci.harvard.edu

Supplementary Information:

Supplementary Methods

Supplementary Tables

Supplementary Table 1. Adapted GRIPP2 short form.

Supplementary Table 2. Angiosarcoma Project patient-reported data summary (N=227 patients).

Supplementary Table 3. Patient-reported sites of primary angiosarcoma (N=227 patients).

Supplementary Table 4. Data elements abstracted from medical records of sequenced Angiosarcoma Project participants (N=36 patients).

Supplementary Table 5. *PIK3CA* mutations observed in angiosarcoma patients

Supplementary Table 6. Information on angiosarcoma patients receiving immunotherapy (N=6 patients).

Supplementary Table 7. Detailed timeline of all treatments received by two head, neck, face, scalp angiosarcoma patients who each had a complete response to Pembrolizumab.

Supplementary Table 8. Angiosarcoma Project forms

Supplementary Methods:

The Angiosarcoma Project (ASCproject) is ongoing and continues to enroll patients. The analyses conducted for this manuscript were performed using information and samples from patients who consented between January 1, 2017 and September 30, 2018.

Patient and public involvement:

Patient and public involvement (PPI) prior to and during the Angiosarcoma Project is described below and has been mapped to the GRIPP2-SF reporting guidelines checklist[4] in Supplemental Table 1 and Extended Data Figures 1-3.

The Angiosarcoma Project and the study website (ASCproject.org) were developed and designed through extensive collaboration with patients. The initial step was to engage patients with angiosarcoma and other members of the angiosarcoma community. Next, an open Facebook Group called The Angiosarcoma Project Working Group was developed in order to provide anybody impacted by the disease (including but not limited to patients, caregivers, loved ones, doctors, etc.) a platform through which the angiosarcoma community could provide feedback on the look, design and specific messaging of the study web portal, and the planned patient intake-survey. In addition, patients in the community were offered other avenues (including telephone, email, and webinars) to connect with study staff if they had any questions or did not want to be part of a social media-based group. Members of the angiosarcoma community that contributed to the project design and development comprised the Angiosarcoma project advisory committee (PAC). We employed an iterative feedback loop where ideas of images and text were presented to the PAC, feedback was collected, and iterations reviewed until there was broad consensus on

all aspects of study design. This resulted in a study website that was unique to the angiosarcoma community, including: the adoption of orange and yellow as the primary colors, the decision to have a turtle prominently displayed as a project mascot, and the use of a silhouetted home page image that reflected the core principles of teamwork and support. In addition, PAC feedback was used to develop answers on the Frequently Asked Questions (FAQ) page around specific issues that reflected the unique concerns brought forth by this community. The PAC also co-developed the survey questions and specific wording therein that would be deployed as part of the study. The PAC also helped develop the language for publicly available data element names and descriptions found for the ASCproject on cBioPortal.org.

The Angiosarcoma Project Working Group on Facebook was also used to engage members of the angiosarcoma community to provide their perspective on other aspects of the study regarding frequency and content of messaging to prospective and current patient participants. Current participants utilized core messaging developed together along with personal messaging to post and share information about the project. Strong relationships between researchers and public and patient representatives has been previously reported to be the most critical element for effective patient-partnered research[5]. Indeed, we found that developing and maintaining trusting relationships and bidirectional channels of open communication between the angiosarcoma community and the study staff have been essential to successfully building and growing this project in a way that has reflected the voice of the angiosarcoma community.

Consented patients participated through a series of steps: the completion of the 17-question intake survey, completion of the medical release form which enabled study staff to acquire

copies of medical records, and the receipt and return of blood and/or saliva specimen kits. Details of this process are outlined elsewhere in this Supplement as well as in Extended Data Figure 5.

From the first day of the ASCproject launch, aggregated and de-identified patient-reported information about the cohort was disseminated back to the community through social media and email. Examples include information such as the number of participants who joined at a given time, and aggregated information regarding anatomic locations of angiosarcoma at diagnosis or age at diagnosis, etc. Video webinars of publicly shared data from the project and live-streamed scientific poster presentation videos were available to participants and also archived to enable on-demand access. Throughout the Angiosarcoma Project, people in the community have always been actively encouraged to provide feedback and ask questions for further clarity, which are answered by study staff within 24 hours. Additionally, regular communications and study updates are shared to patient participants through email, as well as to an even larger public audience via online social media channels.

The Angiosarcoma Project Website:

This research project was designed through extensive collaboration with patients in an online angiosarcoma (AS) support group (The Angiosarcoma Project Working Group in Facebook) who generously provided iterative feedback. Working together with patients, a website was developed (ASCproject.org) that enables AS patients across the United States and Canada to learn about the project, register to participate remotely in this research study, sign an electronic informed

consent (Supplemental Table 8), and provide information about themselves and their cancer. The Angiosarcoma Project website was publicly launched in March 2017. Prior to the date of the public launch, 15 angiosarcoma patients served as beta-testers of the website. All data collected online as part of the Angiosarcoma Project are stored in a secure database.

Patient Registration:

Patients registered for the Angiosarcoma Project by providing their first and last name, email address, and confirmation of a diagnosis of angiosarcoma in a primary screening questionnaire. Registered patients were asked to complete a 17-question survey (all questions optional) about their experiences with angiosarcoma (Supplemental Table 8). Registrants acknowledged that their responses will be stored in a secure database. Included in this acknowledgement statement is an understanding that patients may be re-contacted and that they can withdraw from the study at any time by submitting a written electronic request. Withdrawing from the project stops the generation of any new data from a participant's samples, medical records, or additional surveys, but data already generated will not be deleted.

Patients were able to direct questions to study staff at any time (before, during, or after registration or consent) using the contact information (phone number and email) listed on the ASCproject website. Study updates, which included information about the cohort in aggregate, were shared regularly with registered patients through email communications and online social media channels.

This project also has an entry point for family members or loved ones to fill out a survey about the experiences that a patient had before passing away from angiosarcoma.

Consent:

All patients provided written informed consent electronically for this research study as approved by the Dana-Farber/Harvard Cancer Center Institutional Review Board (DF/HCC Protocol 15-057B).

Registrants who submitted the intake survey were asked to electronically sign an informed consent document for formal enrollment in the study (Supplemental Table 8). Email reminders were sent weekly for three weeks to patients who registered but who had not yet completed the consent process. Signing the consent form allowed study staff to obtain medical records and to send the patient a saliva kit. Within the consent form, patients also could elect to share a blood sample and a portion of their stored tumor samples. After providing informed consent, patients were then asked to complete a medical release form in order to give their full contact information, as well as a list of physicians and institutions that provided their clinical care for angiosarcoma. Email reminders were sent to consented patients who had not completed the medical release form weekly for three weeks. Upon completion of the consent or medical release forms, signed copies were sent electronically to the enrolled patients for their records.

Blood and saliva sample acquisition:

Enrolled patients, who provided informed consent and also completed the medical release form, were mailed saliva kits, if the participant reported a valid residential address in the U.S. or Canada. Prior to shipment, saliva kits were labeled with a unique two-dimensional barcode and a pre-paid business reply label addressed to the Broad Institute Genomics Platform (Cambridge, MA). Each barcode identifier was assigned to a unique participant prior to shipment. Participants provided at least 2 mL of saliva in the included DNA Genotek (Ottawa, Canada) Oragene Discover (OGR-600) tube[1] following the included instructions (see link on ascproject.org/data-release) and returned the kits by mail free of charge. Saliva kits received back at the Broad Institute were logged by their unique barcodes and stored at room temperature until advancement to whole exome sequencing (WES).

Blood kits were mailed to registrants who provided informed consent, opted in to this component of the study, and reported a valid residential address in the U.S. or Canada. Blood kits containing an empty 10 mL Streck (La Vista, NE) Cell-Free DNA BCT tube[2,3] were labeled with a unique two-dimensional barcode and included a prepaid FedEx ClinPak envelope addressed to the Broad Institute Genomics Platform (Cambridge, MA) prior to shipment. Each barcode identifier was assigned to a unique participant prior to shipment. Participants provided a blood sample by bringing the kit and instructions (see link on ascproject.org/data-release) with them to their next regularly scheduled clinical appointment and requesting a courtesy draw. If a courtesy draw was not possible, patients were given the option to go to any Quest Diagnostics™ facility in the U.S. with a voucher. This voucher was for a complimentary blood draw for the patient, with the cost covered by the research project. Blood samples received at the Broad Institute were logged by

their unique barcodes and fractionated. The resultant plasma and buffy coats were stored at -80°C. Buffy coats were used to extract germline DNA for WES if no saliva sample was available.

Medical record and tissue sample acquisition:

Study staff called the offices of the hospitals and physicians listed in each participant's medical release form to confirm the fax number for the medical records department. A detailed request was electronically faxed to each facility that asked for medical records (MR) including clinic notes from treating providers, angiosarcoma treatment data (including radiation and chemotherapy), pathology reports, operative reports, referrals, MD to MD exchange, and genetic testing reports from the date of primary diagnosis through to the date of the request (see Supplemental Table 8 for request form). MR that had not been received after several months were re-requested in the same manner. MR were received by fax, mail, or secure electronic message. All MR were saved to a secure drive to facilitate abstraction. For patients who opted to share portions of their stored tumor samples, the date of each procedure (e.g. biopsy, resection, mastectomy, etc.), the type of procedure, histology, and facility that performed the procedure were abstracted from all pathology reports in order to determine tissue to be requested. Unless the most recent biopsy sample for a given patient measured more than 1 cm, the most recent biopsy samples were not requested in order to avoid exhausting samples that could be needed for future clinical care.

The study staff called the pathology departments associated with each tissue sample to confirm the fax number for tissue requests. A form was faxed to each pathology department requesting 1

hematoxylin and eosin stained slide as well as either 5-micron unstained slides (a minimum of 8 and a maximum of 20 slides) or 1 formalin-fixed paraffin-embedded tissue block. Requests explicitly stated that no sample should be exhausted in order to fulfill the request.

Tissue samples were received at the Broad Institute (Cambridge, MA) by mail. Tissue samples received as blocks were labeled with unique numerical identifiers and sent to the Dana-Farber/Harvard Cancer Center Specialized Histopathology Services - Longwood (SHL) Core, which cut three 30-micron scrolls per block, as well as three unstained slides and one hematoxylin and eosin stained slide for pathology review. SHL was instructed to only cut scrolls if confident that doing so would not exhaust the tissue in the block. The scrolls were then labeled with unique barcode identifiers. Tissue samples received as unstained slides were logged and labeled with unique barcode identifiers.

A hematoxylin and eosin stained slide and three unstained slides from each tumor sample received were sent for expert centralized pathology review to confirm the presence of angiosarcoma in each sample. In cases without a vasoformative component (i.e., composed entirely of cytologically malignant epithelioid and/or spindle cells without vascular channel formation), immunohistochemistry for ERG was performed to confirm endothelial differentiation. Genomic analysis and inclusion of a sample in a publicly released genomic dataset occurred only for samples confirmed to be angiosarcoma. If a given tumor sample was not confirmed to be angiosarcoma, that sample was not included in genomic analysis, but the associated patient information was retained and included in the patient-reported dataset of the ASCproject.

Scrolls and unstained slides were submitted to the Broad Institute Genomics Platform for whole exome sequencing. Germline DNA obtained from saliva or buffy coat samples were sent for WES at the same time as its matched tumor sample. Tumor and germline samples that were not matched were not sent for WES.

Patient-Reported Data:

Patient-reported data (PRD) comes from information provided by patients in the 17-question intake survey (Supplemental Table 8), initially completed by patients at the time of project registration. Patients were given the option to update the survey at any time, but were not explicitly asked or required to provide updates after initial survey completion. All survey questions were optional. Any survey question left blank was categorized as "Not Reported" or "No Response". If survey questions included an option to respond by selecting "I don't know" any response of this option was categorized as "Don't Know" ("DK"). If during the completion of the intake survey, any respondent selected "No" to a question that was set up to trigger subsequent survey question(s) that seek more details, the answer to those subsequent questions was marked as "Not Applicable". The Angiosarcoma Project PRD dataset includes information from all consented patients in the United States and Canada, regardless of the availability of samples or medical records.

In order to protect confidentiality before any data was shared publicly, demographic categories, such as race, that were reported fewer than three times in the dataset were reclassified as "other".

When applicable, data elements from PRD were cleaned to standardize formatting and protect patient confidentiality. For questions in which responses indicated a location in the body, 11 anatomical locations were provided as answer choices (Supplemental Table 8), as well as an option to select "Other" and provide a free text response. Any free text responses provided by patients that did indicate a location in the body from 1 of the 11 provided anatomical locations were classified into that category as part of the cleaning process. Those answers that did not fall into one of the categories or had no subsequent free text were classified as "Other". Patient responses to their referral sources to the ASCproject (Extended Data Figure 10) were manually reviewed and assigned into the following categories "Social Media", "Internet Search/Website", "Doctor", "Friend/Family", "Advocacy Partner", and "Other". Additional information regarding cleaning of PRD is available upon request.

Medical Record Abstraction:

For each consented patient, medical records were requested from all doctors and institutions based in the U.S. and Canada that were listed by the patient in their medical release form. If records were received electronically, they were stored in a secure drive by study staff. Records received in a hard-copy paper format were scanned and securely stored in the same secure drive. To help facilitate manual abstraction, scanned medical records were converted to searchable PDF files by using the Optical Character Recognition (OCR) engine known as Tesseract (LSTM model inside Tesseract version 4.0; (https://github.com/tesseract-ocr/tesseract)). Using the medical records received, 40 pre-determined clinical fields were manually abstracted for each

patient. These clinical fields are listed in detail in a clinical data model that was developed with multiple sarcoma experts (available upon request).

To ensure data integrity, these 40 fields from each medical record were independently abstracted by two different trained abstractors on the study staff. Quality control (QC) for concordance among abstracted data elements was performed by a third trained abstractor. During the QC process, the third abstractor compared all abstracted fields from the other two abstractors. For any clinical fields that lacked 100% concordance, the QC abstractor went back to the records to review. If the QC abstractor was unable to reconcile differences, these fields received additional review from physicians that are experts in angiosarcoma.

When noted in the medical records, dates of diagnoses and treatments were abstracted to the greatest level of detail available in the record. Dates reported in the medical record only as a year were abstracted as the first of the year. Dates reported in the medical record only as a month and year were abstracted as the first of the month. Dates that could be inferred based on notes were indicated as estimated and not directly abstracted. In order to protect patient confidentiality, all dates reported were based on elapsed time relative to the date of primary diagnosis. Ages were grouped into bins for public data releases on cBioPortal.org.

Subclassifications of angiosarcoma (AS):

For patients whose tumor samples underwent sequencing, their first angiosarcoma diagnosis was categorized into eight AS subclassifications using information abstracted from medical records.

These eight subclassifications of AS were defined based on guidance provided by angiosarcoma experts (abdominal area, bladder, cardiac, cutaneous radiation-associated, HNFS (head, neck, face, scalp), lung, primary breast, and spleen).

To assign the cutaneous radiation-associated AS (C-RAAS) subclassification, patients' medical records needed to explicitly indicate that radiation was previously administered (during the treatment of a prior cancer) to the proximal area later affected by angiosarcoma. Additionally for the C-RAAS subclassification, these patients' pathology reports were used to confirm cutaneous AS. For the HNFS AS subclassification, all of the patients' records were reviewed and this reconfirmed that all had cutaneous lesions. The primary breast subclassification included AS patients with breast AS that did not receive prior radiation therapy to the proximal area of the breast.

Attrition of genomic and clinical data in collection and processing:

Of the 223 consented patient participants that were sent either a blood or saliva kit, 180 study participants (81%) returned kits with saliva and/or blood samples, which allowed for extraction of germline DNA.

At the time of this analysis, we had requested medical records for 186 participants from their medical providers and received medical records for 139 patients (75%). Of the 139 participants for whom medical records had been received, we had requested 124 archival tumor samples for 78 participants and received 103 samples for 68 participants (83%). For the other 61 of the 139

participants for whom medical records had been received, these patients' records either did not contain sufficient information to request tissue or showed too little tissue to request for research.

For the 103 tumor samples received, 5 of these samples did not have available matched normal DNA (from blood or saliva) and were not initiated for sequencing. Of the 98 matched tumor/normal sample pairs that were available to initiate DNA extraction, whole exome sequencing was completed on 70 tumor samples from 53 patients (71%, 70/98 sample pairs). For the other 28 tumor samples, 16 samples had insufficient material for sequencing and 12 sets of samples were still in the sequencing pipeline. Of the 70 sequenced tumor samples, 49 samples had at least 10% tumor purity (70%). After being subjected to centralized pathology review, 47 tumor samples were reconfirmed to be angiosarcoma. These 47 tumor samples from 36 angiosarcoma patients were used in subsequent downstream genomic analysis. The resulting analyzed sequence data comprised the Angiosarcoma Project September 2018 dataset that was released on cBioPortal.org along with associated patient-reported and clinical data.

Additional details regarding attrition that occurred at each step of the Angiosarcoma Project as of September 30, 2018 are outlined in Extended Data Figure 5.

Biological Sample Processing and Sequencing:

DNA Isolation

DNA was extracted via the Chemagic MSM I with the Chemagic DNA Blood Kit-96 from Perkin Elmer. This kit combines a chemical and mechanical lysis with magnetic bead-based

purification. Saliva samples were incubated at 50°C for 2 hours. The saliva was then transferred to a deep well plate placed on the Chemagic MSM I. Whole blood samples were incubated at 37°C for 5-10 minutes to thaw. The blood was transferred to a deep well plate with protease and placed on the Chemagic MSM I. The following steps were automated on the MSM I. M-PVA Magnetic Beads were added to the saliva or the blood and protease solution. Lysis buffer was added to the solution and mixed. The bead-bound DNA was then removed from solution via a 96-rod magnetic head and washed in three Ethanol-based wash buffers to eliminate cell debris and protein residue. The beads were then washed in a final water wash buffer. Finally, the beads were dipped in elution buffer to resuspend the DNA sample in solution. The beads were then removed from the solution, leaving purified DNA eluate. DNA samples were quantified using a fluorescence-based PicoGreen assay. For formalin-fixed paraffin-embedded tumor tissues, DNA and RNA were extracted simultaneously using Qiagen's AllPrep DNA/RNA FFPE kit.

Library Construction

Library construction was performed as described[6], with the following modifications: initial genomic DNA input into shearing was reduced from 3μg to 10-100ng in 50μL of solution. For adapter ligation, Illumina paired end adapters were replaced with palindromic forked adapters, purchased from Integrated DNA Technologies, with unique dual-indexed molecular barcode sequences to facilitate downstream pooling. Kapa HyperPrep reagents in 96-reaction kit format were used for end repair/A-tailing, adapter ligation, and library enrichment PCR. In addition, during the post-enrichment SPRI cleanup, elution volume was reduced to 30μL to maximize library concentration, and a vortexing step was added to maximize the amount of template eluted.

In-solution hybrid selection

After library construction, hybridization and capture were performed using the relevant components of Illumina's TruSeq Rapid Exome Kit and following the manufacturer's suggested protocol, with the following exceptions: first, all libraries within a library construction plate were pooled prior to hybridization. Second, the Midi plate from Illumina's TruSeq Rapid Exome Kit was replaced with a skirted PCR plate to facilitate automation. All hybridization and capture steps were automated on the Agilent Bravo liquid handling system.

Preparation of libraries for cluster amplification and sequencing

After post-capture enrichment, library pools were quantified using qPCR (automated assay on the Agilent Bravo), using a kit purchased from KAPA Biosystems with probes specific to the ends of the adapters. Based on qPCR quantification, libraries were normalized to 2nM, then denatured using 0.2N NaOH on the Hamilton Starlet. After denaturation, libraries were diluted to 20pM using hybridization buffer purchased from Illumina.

Cluster amplification and sequencing

Cluster amplification of denatured templates was performed according to the manufacturer's protocol (Illumina) using HiSeq 4000 cluster chemistry and HiSeq 4000 flowcells. Flowcells were sequenced on v1 Sequencing-by-Synthesis chemistry for HiSeq 4000 flowcells. The flowcells were then analyzed using RTA v.1.18.64 or later. Each pool of whole exome libraries was run on paired 76bp runs, reading the dual-indexed sequences to identify molecular indices and sequenced across the number of lanes needed to meet coverage for all libraries in the pool.

Sequencing Data Analysis:

Please contact data@ascproject.org for additional details regarding pipeline information.

Sequence data processing and quality control

Whole exome sequences were captured using Illumina technology and the sequence data processing and analysis was performed using the Picard and Firehose pipelines at the Broad Institute. The Picard pipeline (http://picard.sourceforge.net) was used to produce a BAM file with aligned reads. This includes alignment to the GRCh37 human reference sequence using the BWA aligner[7] and estimation and recalibration of base quality score with the Genome Analysis Toolkit (GATK)[8]. All sample pairs passed through the Firehose pipeline were subjected to QC testing to test for any tumor/normal and inter-individual contamination as previously described.[9,10]

Somatic alterations assessment

The MuTect algorithm was used to identify somatic mutations[10]. To reduce false positive calls, additional analysis of reads covering sites of a putative somatic mutation was done and were realigned with NovoAlign (www.novocraft.com). An additional iteration of MuTect inference on newly aligned BAM files was performed. Furthermore, false-positive somatic mutation calls were filtered using a panel of normals (PoN), generated from the TCGA dataset and oxoG filter[11]. Small somatic insertions and deletions were detected using Strelka algorithm[12], MuTect 2 and SvABA (https://github.com/walaj/svaba). Insertions and deletions which were called by two out of the three methods mentioned above were used for analysis. Somatic mutations including

single-nucleotide variants, insertions, and deletions were annotated using Oncotator[13]. The germline somatic variants were analyzed using the HaplotypeCaller module of GATK[8].

Mutation Significance Analysis

Significance of identified somatic mutations was analyzed using MutSig2CV[14], which uses patient and gene- specific mutation rates to estimate a background model of predicted mutation incidence across the genome. MutSig2CV then factors in biological co-variates such as replication timing and gene-expression level on a gene-by-gene basis to account for the increased mutational rate of certain classes of genes.

Copy number alterations analysis

ReCapseg was used to analyze somatic copy number alterations (SCNA) from whole exome data. ReCapseg assesses homolog-specific copy ratios from segmental estimates of multipoint allelic copy ratios at heterozygous loci incorporating the statistical phasing software (BEAGLE) and population haplotype panels (HAPMAP3)[15,16]. For copy number alteration significance analysis, segmented copy number data was analyzed by GISTIC 2.0, to identify significantly recurring focal and arm-level amplification/deletion peaks[17].

Purity and ploidy estimation

Allele-specific SCNAs and tumor ploidy/purity status were assessed using ABSOLUTE[18]. Angiosarcoma tumor samples which had a purity of 10% or more were submitted to cBioPortal for Cancer Genomics (http://www.cbioportal.org/index.do) and were used for downstream analysis.

Assessment of Tumor Mutation Burden (TMB)

Tumor mutation burden (mutations per megabase) was calculated as the total number of mutations (non-synonymous + synonymous) detected for a given sample divided by the length of the total genomic target region captured with the whole exome sequencing[19]. Samples with TMB ≥ 10 mutations per megabase were classified as hypermutated[20].

Mutational Signature Analysis

Mutational signatures were analyzed using Signature Analyzer tool[21]. Contributions of different mutation signatures were identified for each sample according to distribution of the 6 substitution classes and the adjacent bases immediately 5′ and 3′ of the mutated base, producing 96 possible mutation subtypes. The extracted signatures were compared against the known and validated signatures[22]. A sample was determined to have a dominant signature based on the maximum signature score attributable to that sample. DeconstructSig[23] was used for additional validation of these identified signatures on an individual tumor sample level.

***PIK3CA* Analysis:**

PIK3CA Structural Analysis

Structural analysis was performed using PyMOL (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.). To map the mutations in *PIK3CA* that were observed in breast angiosarcoma samples, the existing structure of the p110alpha protein (PDB ID: 3HHM; https://files.rcsb.org/pub/pdb/validation_reports/hh/3hhm/3hhm_full_validation.pdf) was used.

Dependency data analysis for PIK3CA mutations

CRISPR knockout (KO) dependency data (Avana dataset), cancer cell line mutation calls, and associated cell line and mutation annotations were taken from the DepMap 19Q1 data release (https://depmap.org/portal/download/). The Avana dataset contains gene dependencies estimated for each gene and cell line using the CERES algorithm[24]. CRISPR KO gene dependency scores, which measures the sensitivity to the impact of CRISPR knockout-induced loss of a gene on cell viability, are normalized such that a value of 0 represents the median dependency score of negative control genes and −1 represents the median dependency score of sgRNAs that target pan-essential genes. CRISPR KO gene dependency scores were compared for cell lines with wild-type *PIK3CA* to cell lines harboring hotspot *PIK3CA* mutations and to cell lines with *PIK3CA* mutations observed in this AS cohort. *PIK3CA* hotspot mutations (as defined by depmap.org) were inferred using their frequency of occurrence in TCGA and COSMIC databases.

PIK3CA mutations in GENIE dataset

The GENIE version 5 dataset was used to look for *PIK3CA* mutations and their occurrences in all cancer types[25] (http://genie.cbioportal.org/).

Statistical Analyses:

A two-sided Wilcoxon's rank sum test was used to calculate significance for comparison of TMB across various AS subclassifications. A two-sided Fisher's exact test was used to calculate

significance for univariate frequency comparisons. For dependency data analysis of *PIK3CA* mutations, statistical comparisons were performed using a two-sample, two-sided unpaired t-test. A p-value of < 0.05 was considered to be statistically significant. All statistical analysis was performed using R (version 3.5.2).

Data sharing:

The clinically annotated genomic dataset of the Angiosarcoma Project is shared widely in order for all researchers to be able to utilize this data to better understand angiosarcoma. De-identified data has been shared on a recurring basis as the data is generated. Data from the Angiosarcoma Project is currently online at www.cBioPortal.org, and will be shared in additional data repositories as well in the future. The ASCproject has been registered as a study at dbGaP under the accession number phs001931. For additional details about the Angiosarcoma Project or its data, please contact the study team (data@ascproject.org).

References:

1. Ng, D. P. K., Koh, D., Choo, S. & Chia, K.-S. Saliva as a viable alternative source of human genomic DNA in genetic epidemiology. *Clin. Chim. Acta* **367**, 81–85 (2006).
2. Medina Diaz, I. *et al*. Performance of Streck cfDNA Blood Collection Tubes for Liquid Biopsy Testing. *PLoS One* **11**, e0166354 (2016).
3. Kang, Q. *et al*. Comparative analysis of circulating tumor DNA stability In K3EDTA, Streck, and CellSave blood collection tubes. *Clinical Biochemistry* **49**, 1354–1360 (2016).
4. Staniszewska, S. *et al*. GRIPP2 reporting checklists: tools to improve reporting of patient and public

involvement in research. *Res Involv Engagem* **3**, 13 (2017).

5. Wilson, P. *et al*. ReseArch with Patient and Public invOlvement: a RealisT evaluation – the RAPPORT study. *Health Services and Delivery Research* **3**, 1–176 (2015).
6. Fisher, S. *et al*. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biol*. **12**, R1 (2011).
7. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
8. McKenna, A. *et al*. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**, 1297–1303 (2010).
9. Berger, M. F. *et al*. The genomic complexity of primary human prostate cancer. *Nature* **470**, 214–220 (2011).
10. Cibulskis, K. *et al*. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnology* **31**, 213–219 (2013).
11. Costello, M. *et al*. Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Res*. **41**, e67 (2013).
12. Saunders, C. T. *et al*. Strelka: accurate somatic small-variant calling from sequenced tumor–normal sample pairs. *Bioinformatics* **28**, 1811–1817 (2012).
13. Ramos, A. H. *et al*. Oncotator: cancer variant annotation tool. *Hum. Mutat*. **36**, E2423–9 (2015).
14. Lawrence, M. S. *et al*. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
15. Browning, B. L. & Yu, Z. Simultaneous Genotype Calling and Haplotype Phasing Improves Genotype Accuracy and Reduces False-Positive Associations for Genome-wide Association Studies. *The American Journal of Human Genetics* **85**, 847–861 (2009).
16. Consortium, T. I. H. 3. & The International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).

17. Beroukhim, R. *et al*. Assessing the significance of chromosomal aberrations in cancer: Methodology and application to glioma. *Proceedings of the National Academy of Sciences* **104**, 20007–20012 (2007).
18. Carter, S. L. *et al*. Absolute quantification of somatic DNA alterations in human cancer. *Nature Biotechnology* **30**, 413–421 (2012).
19. Chalmers, Z. R. *et al*. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med*. **9**, 34 (2017).
20. Campbell, B. B. *et al*. Comprehensive Analysis of Hypermutation in Human Cancer. *Cell* **171**, 1042–1056.e10 (2017).
21. Kim, J. *et al*. Somatic ERCC2 mutations are associated with a distinct genomic signature in urothelial tumors. *Nat. Genet*. **48**, 600–606 (2016).
22. Alexandrov, L. B. *et al*. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
23. Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B. S. & Swanton, C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol*. **17**, 31 (2016).
24. Meyers, R. M. *et al*. Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. *Nature Genetics* **49**, 1779–1784 (2017).
25. Consortium, T. A. P. G. & The AACR Project GENIE Consortium. AACR Project GENIE: Powering Precision Medicine through an International Consortium. *Cancer Discovery* **7**, 818–831 (2017).

Supplementary Table 1. Adapted GRIPP2 Short Form.

Section and Topic	Item
Aim	The Aim of Patient and Public Involvement (PPI) in this project was to engage patients with angiosarcoma remotely across the U.S. and Canada to contribute to the design and deployment of a research study that would empower patients to share their medical information, biospecimens, and experiences with researchers to accelerate discoveries in their own disease.
Methods	**1. Engagement Strategy and Study Design** A. Identified and engaged key members of the angiosarcoma community B. Co-created study website and materials through iterative feedback C. Developed comprehensive community engagement strategy **2. Outreach and Accrual** A. Gathered patients' preferences and perspectives on outreach approaches and messaging B. Partnered with angiosarcoma and pan-sarcoma advocacy groups to raise project awareness and perform outreach **3. Continued Engagement and Project Iteration** A. Disseminated de-identified patient-reported data from the project in aggregate to participants, the angiosarcoma community, and the public B. Responded consistently and rapidly to questions and feedback from the angiosarcoma community C. Regularly shared new data analyses with participants, the angiosarcoma community, and the public D. Shared lay-friendly video walkthroughs of ASCproject posters at scientific conferences in real time E. Shared email newsletter updates on project progress, including distribution of pre-publication manuscript
Results	**1. Engagement Strategy and Study Design (Extended Data Figure 1)** A. Study staff conducted 40 design discussions through social media and 13 website walkthroughs B. Home page images, colors, and messaging all reflect community feedback C. FAQ page addresses unique concerns raised by the community D. Patient intake survey questions co-developed with the community **2. Outreach and Accrual (Extended Data Figure 2)** A. 110 patients registered for the project within the first week, including 54 in the first 8 hours B. Advocacy partners raised awareness for the project through their existing communication networks C. Community members developed a calendar to post regularly about the project in online support groups **3. Continued Engagement and Project Iteration (Extended Data Figure 3)** A. Study staff posted over 110 project updates on social media within the first year B. Five lay-friendly poster video walkthroughs were shared by study staff from scientific conferences C. Study staff had 8 discussions via social media to rework project content and processes D. Publicly available data element names and descriptions were co-developed with the community
Discussion	Working directly with the angiosarcoma community enabled the study staff to build this project in a way that reflected the voices of patients with angiosarcoma. We were able to rapidly accrue patients, with 110 registering within the first week. The ability to generate novel insights into the genomics of this rare disease was contingent upon PPI throughout the design, deployment, accrual, and refinement of the project.
Reflections	The use of PPI in the Angiosarcoma Project was foundational in building a true patient partnership which enabled rapid enrollment, data generation, and discoveries with implications for clinical care. The methods described here may pose additional challenges in larger patient populations spanning many different support groups. Investigators should be sensitive to ensuring parity in terms of engagement opportunities and varied perspectives when working with a large number of groups that may represent a patient community.

Supplementary Table 2. Angiosarcoma Project Patient-Reported Data Summary (N=227 patients)

Question / Response	YES	NO	NOT REPORTED	DON'T KNOW	N/A
Are you currently being treated for your angiosarcoma?	99 (43.6%)	118 (52.0%)	5 (2.2%)	5 (2.2%)	--
Have you had surgery to remove angiosarcoma?	177 (78.0%)	49 (21.6%)	1 (0.4%)	0 (0.0%)	--
If so, did the surgery remove all known cancer tissue (also known as "clean margins")?	142 (62.6%)	30 (13.2%)	1 (0.4%)	5 (2.2%)	49 (21.6%)
Have you had radiation as a treatment for angiosarcoma? If you had radiation for other cancers, we will ask you about that later.	93 (41.0%)	130 (57.3%)	3 (1.3%)	0 (0.0%)	1 (0.4%)
Were you ever diagnosed with any other kind of cancer(s)?	95 (41.9%)	130 (57.3%)	1 (0.4%)	1 (0.4%)	--
Have you had radiation as a treatment for another cancer(s)?	60 (26.4%)	164 (72.2%)	1 (0.4%)	1 (0.4%)	1 (0.4%)
Do you consider yourself Hispanic, Latino/a or Spanish?	9 (4.0%)	214 (94.3%)	4 (1.8%)	0 (0.0%)	--

Supplementary Table 3. Patient-reported sites of primary angiosarcoma (N=227 patients).
The number of patients and the percentage of all patients who selected a given answer is indicated in the right column.

Patient-Reported Site of Primary Angiosarcoma	Number of Patients (%)
Breast (encompassing both cutaneous and parenchymal tumors of the breast)	95 (42%)
Head, Face, Neck, Scalp	56 (25%)
Bone/Limb	21(9%)
Heart	12 (5%)
Abdominal area	6 (3%)
Liver	4 (2%)
Spleen	3 (1%)
Lung	2 (1%)
Brain	1 (0.4%)
Multiple locations	16 (7%)
Other locations	8 (3.5%)
No answer to the question	3 (1%)

Supplementary Table 4. Data elements abstracted from medical records of sequenced Angiosarcoma Project participants (N=36 patients).

Obtained medical records contained sufficient information to assess and abstract all of the above data elements for 35 patients. For one additional patient, records were only sufficient to abstract 13 data elements. When available in medical records, dates of diagnoses and treatments were abstracted. To protect patient confidentiality, all dates reported in public releases are indicated as elapsed time relative to the date of initial angiosarcoma diagnosis. For additional details on medical record abstraction and data element masking, please see Supplementary Methods.

Patient-Level Data Elements	**Sample-Level Data Elements**
Biological Sex	Sample Collection Date
Age at Diagnosis	Procedure Location
Diagnositc Biopsy Location	Procedure Type
Primary Tumor Site	Vasoformative
Medications Received	Epithelioid
Radiation Associated Angiosarcoma	Spindle Cell
Adjuvant Radiation	Nuclear Grade
Presence of Metastatic Disease	
Time to Metastatic Diagnosis	
Metastatic Sites at Metastatic Diagnosis	
Ever Metastatic Sites	
Local Recurrence	
Local Recurrence Sites	
Time to Local Recurrence(s)	
Other Cancer Type	

Supplementary Table 5. *PIK3CA* mutations observed in angiosarcoma patients

PI3K Protein Change	Patients with this mutation	Tumor fraction	Co-occurrence with other *PIK3CA* mutations in a patient	OncoKB Annotation	Cancer Hotspot	3D Hotspot (as defined by Gao *et al.* 2017)	COSMIC occurrences	GENIE.v5 occurrences	Cancer cell lines (using DepMap) with observed mutation	Cancer cell types with functional validation *in vitro*	*In vitro* functional validation references	Other references	**Cumulative evidence of mutation being activating**
M1043I	ASCProject_loC8UQUk	0.09	T957P	Oncogenic, level_3b	Yes	No	107	76		Breast, Colorectal	PMID: 26627007, PMID: 17376864, PMID: 15930273		Strong
M1043V	ASCProject_GXTMTxU7	0.23		Oncogenic, level_3b	Yes	No	107	36 (multiple breast, multiple carcinomas)		Breast	PMID: 26627007, PMID: 17376864		Strong
R88Q	ASCProject_3NfIfGHo	0.23		Likely Oncogenic, level_3b	Yes	Yes	89	202	4 (2 *PIK3CA* mutation - soft tissue, uterus; 2 with other non-conserving *PIK3CA* mutations - colorectal, uterus)	Bone	PMID: 18829572	PMID: 22949682 (Structural analysis)	Strong
N1044K	ASCProject_EAuOu4uD	0.09		Oncogenic, level_3b	Yes	No	32	35 (multiple breast, multiple carcinomas)		Breast	PMID: 26627007		Strong
P124L	ASCProject_blh1uyhK	0.33, 0.07		Unknown, level NA	No	No	2	1 (Breast), 1(Angiosarcoma - P124Q), 1(Oligodendroglia - P124E)	1 (urinary track)	Urothelial carcinoma	PMID: 22430209 PMID: 27465249		Medium
G914R	ASCProject_EdCVC0um	0.06		Likely Oncogenic, level_3b	No	No	1	8 (1 breast, 6 carcinomas)	1(colorectal)		PMID: 22729224, PMID: 27191687, PMID: 28502725		Medium
N107T	ASCProject_RvtOtjtj	0.20, 0.14		Predicted Oncogenic, level_3b	No	No	1	1 (Breast)				PMID: 22949682 (Structural study on mutation in G106V)	Medium
T957P	ASCProject_loC8UQUk	0.10, 0.20	M1043I	Unknown, level NA	No	No	NA	1 (Breast)					Unknown

Supplementary Table 6. Information on angiosarcoma patients receiving immunotherapy (N=6 patients).

Project ID	Primary Site	Tumor Mutation Burden (mutations per megabase)	Immunotherapy	IO Duration (days)	IO Response	Discontinuation	Received Tx After IO	Notes about IO Response
ASCProject_bRSGSlCG	LUNG	0.4	PEMBROLIZUMAB	42	Varied response among lesions	Side Effects	YES	--
ASCProject_QYsAsrTO	PRIMARY BREAST	4.6, 3.3 (two tumor samples)	NIVOLUMAB	177	Progression	Progression	YES	--
			PEMBROLIZUMAB	86	Progression	Side Effects	UNKNOWN	Pembrolizumab was stopped due to adverse effects
ASCProject_XjHAU9uR	CARDIAC	1.4	PEMBROLIZUMAB	169	Progression	Progression	YES	--
ASCProject_NdUxUwCM	HNFS	62.3	PEMBROLIZUMAB	1	Received one dose, progression	Progression, Side Effects	YES	Only 1 dose due to RA, clinical disease progression
ASCProject_KxFGsofW	HNFS	78.5	PEMBROLIZUMAB	219	Complete	Side Effects	NO	Lasting response (NED)
ASCProject_dyhLT8sG	HNFS	138.9	PEMBROLIZUMAB	249; 30	Stable disease	Side Effects	NO	Off therapy due to autoimmune hepatitis

IO: Immunotherapy; Tx: Therapy (non-IO); HNFS: Head, Neck, Face and Scalp; RA: Rheumatoid Arthritis; NED: No Evidence of Disease

Supplementary Table 7. Detailed timeline of all treatments received by two head, neck, face, scalp angiosarcoma patients who each had a complete response to Pembrolizumab.

The number of days indicated below are calculated relative to the initial diagnosis of angiosarcoma (day of diagnosis being day 0).

Patient ID	Start (Days)	Stop (Days)	Drug	Treatment Category
Patient 1				
ASCProject_KxFGsofW	15	139	PACLITAXEL	CHEMOTHERAPY
ASCProject_KxFGsofW	15	148	BEVACIZUMAB	VEGF INHIBITOR
ASCProject_KxFGsofW	149	169	NO TREATMENT	NA
ASCProject_KxFGsofW	170	230	GEMCITABINE	CHEMOTHERAPY
ASCProject_KxFGsofW	170	230	DOCETAXEL	CHEMOTHERAPY
ASCProject_KxFGsofW	231	259	NO TREATMENT	NA
ASCProject_KxFGsofW	260	408	DOXORUBICIN	CHEMOTHERAPY
ASCProject_KxFGsofW	266	428	DEXRAZOXANE	CHEMOPROTECTIVE AGENT
ASCProject_KxFGsofW	300	408	IFOSFAMIDE	CHEMOTHERAPY
ASCProject_KxFGsofW	429	600	NO TREATMENT	NA
ASCProject_KxFGsofW	601	642	CLINICAL TRIAL	CLINICAL TRIAL
ASCProject_KxFGsofW	643	649	NO TREATMENT	NA
ASCProject_KxFGsofW	650	713	CYCLOPHOSPHAMIDE	CHEMOTHERAPY
ASCProject_KxFGsofW	714	747	NO TREATMENT	NA
ASCProject_KxFGsofW	748	814	CLINICAL TRIAL	CLINICAL TRIAL
ASCProject_KxFGsofW	814	833	CLINICAL TRIAL	CLINICAL TRIAL
ASCProject_KxFGsofW	833	860	CRIZOTINIB	TYROSINE KINASE INHIBITOR
ASCProject_KxFGsofW	860	1079	PEMBROLIZUMAB	ANTI-PD-1 MAB
ASCProject_KxFGsofW	1080	1847	NO TREATMENT	NA
Patient 2				
ASCProject_dyhLT8sG	0	450	NO TREATMENT	NA
ASCProject_dyhLT8sG	451	667	PACLITAXEL PROTEIN BOUND	CHEMOTHERAPY
ASCProject_dyhLT8sG	668	786	NO TREATMENT	NA
ASCProject_dyhLT8sG	787	907	CLINICAL TRIAL	CLINICAL TRIAL
ASCProject_dyhLT8sG	908	1133	NO TREATMENT	NA
ASCProject_dyhLT8sG	1134	1194	PACLITAXEL PROTEIN BOUND	CHEMOTHERAPY
ASCProject_dyhLT8sG	1195	1219	NO TREATMENT	NA
ASCProject_dyhLT8sG	1220	1469	PEMBROLIZUMAB	ANTI-PD-1 MAB
ASCProject_dyhLT8sG	1470	1681	NO TREATMENT	NA
ASCProject_dyhLT8sG	1682	1712	PEMBROLIZUMAB	ANTI-PD-1 MAB
ASCProject_dyhLT8sG	1713	2549	NO TREATMENT	NA

NA: Not Applicable, ANTI-PD-1 MAB: Anti-programmed-cell death-1 monoclonal antibody

Supplementary Table 8. Angiosarcoma Project forms.
Supplementary Table 8 contains a blank copy of each of the following Angiosarcoma Project documents and forms: primary screening questionnaire, patient intake survey, consent form, medical release form, and medical record request form.

Provided as an external excel file.