nature research

Corresponding author(s): Nikhil Wagle, MD

Last updated by author(s): 12-5-2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

All software and methods for genomic data generation is described in detail in Supplementary Methods.
All genomics data collection software and tools used in this study are published and/or publicly available. Exome BAM files were produced with the Picard pipeline (http://picard.sourceforge.net).

Data analysis

All software and pipelines for genomic analysis are described in detail in Supplementary Methods.

All genomics analysis software and tools used in this study are published and publicly available. Whole exome genomic analysis was performed with the Firehose environment (http://www.broadinstitute.org/cancer/cga/Firehose) using genomics tools available at http://www.broadinstitute.org/cancer/cga. As detailed in the Supplementary Methods section, these tools include: ConTest, MuTect, Indelocator, SvABA, Recapseg, GISTIC, Strelka, Mutect2, Novoalign, Oncotator, GATK Haplotype Caller, MutSig2CV and Absolute.

--Mutational Signature analysis was performed using the following publicly available tools as described in the Supplementary Methods:
1. Signature Analyzer (https://software.broadinstitute.org/cancer/cga/msp)
2. DeconstructSig (https://github.com/raerose01/deconstructSigs)
--Protein structure visualization analysis was performed using PyMOL (The PyMOL Molecular Graphics System,325 Version 1.2r3pre, Schrödinger, LLC.)
--As described in the Supplementary Methods, the following additional publicly available datasets were used for analysis and validation:
1. Cancer Dependency Map - DepMap 19Q1 data release (https://depmap.org/portal/download/)
2. AACR-GENIE - GENIE version 5 (http://genie.cbioportal.org/)

Additionally, the following software/tools were used to analyze data:
1. Picard (http://picard.sourceforge.net) – version 2.10.10
2. ConTest - GenomeAnalysisTK-3.1
3. BWA Aligner -version 0.5.9
4. Mutect version 1.1.6-0-g6fe4f4c

5. Mutect2 Version=3.4-45-g6e00632
6. GATK Haplotype Caller Version=3.1-1-g07a4bf8
7. NovoAlign - version .novocraft-2.07.18
8. Strelka version 2.0.13
9. SvABA version 0.2.1
10. Indelocator – GenomeAnalysisTK-3.1
11. Oncotator v1.9.1.0
12. MutSig2CV – v3.1
13. ReCapseg (BEAGLE, HAPMAP)- 1.5.0.0
14. GISTIC version 2.0.22
15. Absolute version 1.06
16. Signature Analyzer version 1
17. DeconstructSig version 1.8.0
18. PyMOL Version 1.2r3pre
19. R 3.5.2
20. Tableau 2018.3

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).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

De-identified patient-reported and clinical data, as well as analyzed genomic data generated by the Angiosarcoma Project is currently available at cBioPortal for Cancer Genomics (cBioPortal.org; https://www.cbioportal.org/study/summary?id=angs_project_painter_2018). The study dataset is de-identified and patient ID's are masked before it is shared in order to protect patient confidentiality (see Supplementary Methods for more details). There are no restrictions on data availability. Please include a citation statement to accompany any publication that uses the Angiosarcoma Project dataset: "The results included here include the use of data from The Angiosarcoma Project (https://ascproject.org/), a project of Count Me In (https://joincountmein.org/)."
The ASCproject has been registered as a study at dbGaP under the accession number phs001931.
Contact data@ascproject.org for additional details regarding data generated by the Angiosarcoma Project.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size for this study could not be pre-determined given that this project works with a rare cancer population and implemented a new patient-partnered research approach. As such, assumptions about sample size were not possible. To the best of our knowledge, this is the largest reported cohort of angiosarcoma samples that have undergone whole exome sequencing. The Angiosarcoma Project is an ongoing sequencing study of angiosarcoma. As of September 30, 2018, 227 angiosarcoma patients consented to enroll in the study. We performed whole exome sequencing on the 70 tumor samples paired with a matched normal sample (saliva or blood) that had been obtained by this time. Forty-seven samples from 36 patients were used for subsequent genomic analysis after assessment of sufficient tumor purity (greater than or equal to 10%) and confirmation as angiosarcoma by centralized pathology review (Extended Data Figure 5 details attrition). Apart from these considerations, there were no additional selection criteria for these 47 samples.
Data exclusions	As described above in the 'Sample size' section above, some tumor samples that underwent whole exome sequencing were subsequently excluded from subsequent genomic analysis due to quality-control. For inclusion in subsequent genomic analysis, samples needed to contain at least 10% tumor purity (as estimated by Absolute) and also had to be confirmed as angiosarcoma (through centralized pathology review). This sequencing QC and pathology review criteria were pre-established. Extended Data Figure 5 includes additional details regarding sample attrition.
Replication	No replication was performed in this study.

Randomization	No randomization was used in this study as it is an observational study.
Blinding	No blinding was used in this observational study as all patients by entry criteria had a diagnosis of angiosarcoma.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☐	☒ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	This study did not involve a clinical trial and no intervention was performed as part of this research. Patients' medical records and biological samples (tumor, saliva, and/or blood) were obtained with their informed consent and the approval of the institutional review board, as described in Methods and Supplementary Methods. Through an online web-portal (https://ascproject.org/), this study was open for enrollment to all patients with angiosarcoma living within the United States and Canada. Information regarding enrolled patient demographics, disease characteristics, and clinical treatments is described in the manuscript and and shown in Figures (e.g. Figures 1B,2A-C, 3A, 4A and Extended Data Figures 4, 6, 9) and Supplementary Tables 2, 3, 6, 7. Patient demographics for the overall cohort of 227 consented patients and the 36 patient subset whose samples underwent sequencing include: age at diagnosis with angiosarcoma (mean: 53.1 years) and years from angiosarcoma diagnosis to ASCproject registration (mean: 3.6 years). Additionally, clinical disease characteristics such as sites of current lesions and sites of lesions ever detected are presented.
Recruitment	This project does not directly recruit patients through any recruitment offices or specific clinical institutions. Enrollment for this project is patient-driven and achieved through a web-based portal (https://ascproject.org/). This study is open for any patients with angiosarcoma who live within the United States and Canada. Patients learn about this project from other patients, through word-of-mouth, and via social media, as well as from scientific outreach and presentations. Enrolled patient-reported referral sources to the Angiosarcoma Project are depicted in Extended Data Figure 10. Due to the web-based enrollment of the Angiosarcoma Project, there is a selection for patients with access to the Internet. The language used in the enrollment and consent process is English, which will select for patients comfortable with this language. Although this use of online engagement and patient-driven registration may be predicted to skew the demographics of study subjects toward younger or less sick patients, we found that the average age at AS diagnosis in our cohort was just slightly lower than that of a single institution AS study2 (53 and 62 years, respectively) and that more than a third of patients (86/227) enrolled within one year of their primary AS diagnosis.
Ethics oversight	The study protocol was approved by the Dana-Farber/Harvard Cancer Center Institutional Review Board (DF/HCC Protocol 15-057B).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	This study did not involve a clinical trial.
Study protocol	The study protocol is DF/HCC Protocol 15-057B, as approved by the Dana-Farber/Harvard Cancer Center Institutional Review Board. Additional information regarding study procedures and protocols can be found in Methods and Supplementary Methods of this manuscript.
Data collection	The Angiosarcoma Project is ongoing and was launched in January 2017. This study has web-based project enrollment, consent, and a patient intake survey. This publication analyzes information and samples associated with patients who registered for the Angiosarcoma Project between January 1, 2017 and September 30, 2018. 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.

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 (data dictionary 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.

Outcomes

Outcome measures were not considered in this study.