Original Article

Investigating the genomic landscape of sarcoma in India: Discoveries from a retrospective observational approach

ABSTRACT

Background: Due to the complex histological and genomic nature of sarcomas, diagnosing and treating them has proven challenging. Delving into the genomic profiles and molecular markers linked to different sarcoma subtypes will aid in overcoming these obstacles and identifying new potential therapeutic targets.

Objectives: The primary objective of this study was to investigate the genomic complexity of sarcoma, while the secondary objective was to identify potential therapeutic targets in the patients with sarcoma from India.

Materials and Methods: This retrospective observational study was conducted from January 2020 to February 2024 at 4basecare Precision Health Pvt. Ltd., Bengaluru, India. We carried out comprehensive genomic profiling using gene panels or exome sequencing, including assessment of immunotherapy biomarkers (tumor mutation burden (TMB), microsatellite instability (MSI), programmed death-Ligand 1 (*PD-L1*)), in a cohort of 263 patients with sarcoma, categorized into 25 sarcoma types, for the present retrospective analysis.

Results: We included 263 patients with sarcoma in our study and identified a diverse landscape of pathogenic variants across 138 genes, in 69.5% (183 patients) of the cohort. SNVs were prevalent in *TP53* (25.1%; 66 patients), *KIT* (5.7%; 15 patients), *PTEN* (4.6%; 12 patients), and *RB1* (4.6%; 12 patients), while *CDK4* (5.2%; 17 patients) and *MDM2* (5.7%; 15 patients) gene amplifications and *SS18-SSX2* (1.1%; 3 patients), *EWSR1-FLI1* (0.8%; 2 patients), and *ASPSCR1-TFE3* (0.8%; 2 patients) gene fusions were recurrent. The majority of the patients harbored mutations affecting cell cycle control (39.2%; 103 patients), *PI3K/AKT/MTOR* (17.9%; 47 patients), and *RAS/RAF/MAPK* (14.8%; 39 patients) pathways. The average TMB was 7 mutations/mb, with 13.3% (35 patients) classified as TMB-H. Around 59.3% of the cohort (156 patients) harbored clinically actionable variants of therapeutic significance, including 8.7% of the cohort (23 patients) who were eligible for FDA/NCCN approved therapies.

Conclusion: The findings emphasize the clinical usefulness of genomic profiling in guiding precision medicine for sarcoma treatment. Our research offers valuable insights into the genetic makeup of sarcomas, serving as a basis for devising efficient and precise diagnostic approaches and for planning preclinical and clinical studies to develop innovative treatment strategies.

Keywords: Comprehensive genomic profiling (CGP), genomic landscape, precision oncology, sarcoma

Pramod S Chinder, R Veena[1], Mohamed Zehran[2], Amit Rauthan[3], Anant Pai[4], Krishna Prasad[5], Ravindra Vottery[6], Mangesh Kamath[7], Aju Mathew[8], Aparna Sreevatsa[9], C B Avinash[10], Dinesh Shet[11], Favaz Ali[12], P S Dattatreya[13], Meenu Walia[14], Vineet G Gupta[15], Bhuvan Chugh[16], Sajjan Rajpurohit[17], Lalit M Sharma[18], Amit Kumar[19], Deepak Mulajkar[20], Ajaykumar Singh[21], Neemesh Lodh[22], Nilesh Lokeshwar[23], Rushabh Kothari[24], Abhinav Zawar[25], Amit Ghanekar[26], Anup Toshniwal[27], Imran Shaikh[28], Jay Anam[29], Mansi Shah[30], Mubarakunnisa Tonse[31], Muzammil Shaikh[32], R K Deshpande[31], Sewanti Limaye[33], Uma Dangi[34], Vaibhav Amale[35], Vijay Patil[36], Akshay Shah[37], Kumar Prabhash[21], Vijay Sharnagat[38], Preetam Jain[39], Pushpak Chirmade[40], Sachin Almel[36], Sunil Chopade[41], Paridhy V Subramanyam[42], Sreekanthreddy Peddagangannagari[42], Giridharan Periyasamy[42], Kshitij D Rishi[42], Hitesh M Goswami[42], Vidya H Veldore[42]

The Yellow Ribbon, Bengaluru, Karnataka, [1]Triesta Sciences, A Unit of HCG, Bengaluru, Karnataka, [2]Apollo Speciality

Access this article online	
Website: https://journals.lww.com/crst	Quick Response Code
DOI: 10.4103/crst.crst_211_24	

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: WKHLRPMedknow_reprints@wolterskluwer.com

How to cite this article: Chinder PS, Veena R, Zehran M, Rauthan A, Pai A, Prasad K, *et al.* Investigating the genomic landscape of sarcoma in India: Discoveries from a retrospective observational approach. Cancer Res Stat Treat 2025;8:173-83.

Submitted: 10-Sep-2024 **Revised:** 07-Feb-2025
Accepted: 15-Feb-2025 **Published:** 12-Aug-2025

Hospital, Chennai, Tamil Nadu, [3]Manipal Hospital, Bengaluru, Karnataka, [4]KMC Hospital, Manipal, Karnataka, [5]MIO Hospital, Mangalore, Karnataka, [6]Yashoda Hospital, Secunderabad, Telangana, [7]Healius Cancer and Hematology, Bengaluru, Karnataka, [8]Ernakulam Medical Center, Ernakulam, Kerala, [9]Sahyadri Narayana Hospital, Shimoga, Karnataka, [10]Clearmedi Radiant Hospital, Mysore, Karnataka, [11]Father Muller Medical College Hospital, Mangalore, Karnataka, [12]Lakeshore Hospital, Kozhikode, Kerala, [13]Renova Soumya Cancer Center, Secunderabad, Telangana, [14]Max Super Speciality Hospital, Delhi, [15]Artemis Hospital, Delhi, [16]Max Hospital, Delhi, [17]BLK- Max Super Speciality Hospital, Delhi, [18]Mahatma Gandhi Hospital, Jaipur, Rajasthan, [19]Jayaprabha Medanta Super Speciality Hospital, Patna, Bihar, [20]Army Hospital, Delhi, [21]Tata Memorial Hospital, Homi Bhabha National Institute, Mumbai, Maharashtra, [22]AIMS Hospital, Mumbai, Maharashtra, [23]Lilavati Hospital, Mumbai, Maharashtra, [24]Oncowin Cancer Center, Ahmedabad, Gujarat, [25]Kamalnayan Bajaj Hospital, Aurangabad, Maharashtra, [26]Fortis Hospital, Mumbai, Maharashtra, [27]Government Cancer Hospital, Chhatrapati Sambhajinagar, Maharashtra, [28]Kokilaben Hospital, Mumbai, Maharashtra, [29]Specialty Surgical Oncology Hospital, Mumbai, Maharashtra, [30]HCG Hospital, Ahmedabad, Gujarat, [31]ACI Cumballa Hill Hospital, Mumbai, Maharashtra, [32]Nanavati Max Super Speciality Hospital, Mumbai, Maharashtra, [33]Reliance Foundation Hospital, Mumbai, Maharashtra, [34]Horizon Hospital, Sangli, Maharashtra, [35]H.N Reliance Hospital, Mumbai, Maharashtra, [36]PD Hinduja Hospital, Mumbai, Maharashtra, [37]Holy Spirit Hospital, Mumbai, Maharashtra, [38]Manav Kalyan Kendra, Mumbai, Maharashtra, [39]Bombay Hospital, Mumbai, Maharashtra, [40]Dr. Pushpakchirmade Cancer Center, Mumbai, Maharashtra, [41]Jaslok Hospital, Mumbai, Maharashtra, [42]4basecare Precision Health Pvt Ltd, Bengaluru, Karnataka, India

Address for correspondence: Dr. Vidya H Veldore,
4basecare Precision Health Pvt Ltd., BHIVE Workspace Whitefield, 8th Floor, Whitefield Main Rd, Brigade Metropolis, Garudachar Palya, Mahadevapura, Bengaluru - 560 048, Karnataka, India.
E-mail: vidya@4basecare.com

PUTTING IN PERSPECTIVE

Central question

- What is the genomic landscape of various sarcoma types in the cohort and what is the clinical utility of genomic profiling in guiding precision medicine for sarcoma management?

Key findings

- Comprehensive genomic profiling identified driver/pathogenic mutations in 69.5% of the cohort (183 patients), including 25 sarcoma types.
- In total, 59.3% ($n = 156$) of the patients were found to be eligible for treatment with available therapy, including 8.7% ($n = 23$) patients eligible for FDA/NCCN approved drugs.
- SNVs in *TP53, KIT, PTEN*, and *RB1*; gene amplifications in *CDK4* and *MDM2*; and gene fusions in and *SS18-SSX2, EWSR1-FLI1,* and *ASPSCR1-TFE3* were prevalent.
- The average TMB of the cohort was 7 mutations/mb.
- Cell cycle control, *PI3K/AKT/MTOR*, and *RAS/RAF/MAPK* pathways were frequently affected in the present cohort.

Impact

- Our study provides a comprehensive view of the diverse genomic and therapeutic landscape across 25 sarcoma types, revealing the genomic complexity, heterogeneity, and proportion of patients with targetable biomarkers in the Indian cohort.
- By identifying the affected genomic pathways, this research supports the exploration of therapeutic options and the design of clinical trials to develop potential treatment strategies.

INTRODUCTION

Sarcomas are a rare, heterogeneous group of mesenchymal malignancies including soft tissue and bone tumors.[1] The incidence of sarcoma in India and South Asian countries is less than 3 per 100,000.[1] More than 100 types of sarcomas have been recognized.[2] Owing to the complex histology and genomic heterogeneity, diagnosis and disease management have been challenging.[3] *En bloc* surgical resection for localized tumors followed by radiation therapy and chemotherapy/neoadjuvant therapy remain the standard-of-care for sarcoma; while curative resection is not an option in metastatic disease, systemic therapy is relied upon, which leads to poor prognosis and patient outcomes.[4] About 40-50% of patients with sarcoma progress to metastatic disease with limited therapeutic options. In metastatic sarcomas, the

median overall survival (OS) is 12-20 months on palliative chemotherapy.[5]

Hallmark genomic alterations (single nucleotide variants (SNVs), copy number variants (CNVs), and gene fusions/chromosomal translocations) have been detected in various sarcomas.[6] Sarcomas lacking these markers harbor numerous nonspecific alterations. Next-generation sequencing (NGS) has been reported to improve accuracy of diagnosis and provide clinical benefits with targeted therapy in sarcomas.[6-8] Due to its relative rarity and diversity among sarcoma subtypes, as well as within these subtypes, there remains a limited understanding of the sarcoma genomic landscape. A comprehensive approach including accelerated accurate diagnosis, identifying targets for therapy, and an in-depth understanding of the genomic profile of sarcomas is essential for efficient disease management, leading to improved patient outcomes.

Here, we aim to explore the genomic landscape of sarcoma in an Indian cohort. Our study investigates the prevalence of clinically actionable variants with therapeutic significance, across different sarcoma types. This is the first comprehensive NGS-based study in India to explore the genomic and molecular complexities of sarcoma.

MATERIALS AND METHODS

General study details

The present retrospective observational study was conducted from January 2020 to February 2024 at 4basecare Precision Health Pvt. Ltd., Bengaluru, India, supported by internal funding. This study, approved by an independent ethical committee and review board (Jehangir Clinical Development Center (JCDC), India), was carried out in accordance with principles of Declaration of Helsinki, Indian Council of Medical Research (ICMR), and Good Clinical Practice Guidelines. The study protocol was approved by the ethical committee [Supplementary Appendix 1]. Written informed consent was obtained from the study participants for research publications with deidentified data. This study was not included in any clinical trial.

Participants

In this study cohort, we included 263 patients diagnosed with various types of sarcomas. These were walk-in patients who were referred for genomic profiling after lines of standard-of-care therapy and/or had disease progression during the lines of therapy. A clinical diagnosis of various sarcoma subtypes, carcinosarcoma and sarcomatoid carcinoma, with or without prior treatment, was considered as the inclusion criterion; patients with cancer types other than the above mentioned were excluded from the study. The study participants were randomly chosen for this retrospective study, irrespective of age and gender.

Aims/objectives

The primary objective of this study was to analyze the genomic complexity of sarcoma by assessing genetic alterations, mutational burden, and molecular subtypes. The secondary objective was to identify potential therapeutic targets in Indian patients with sarcoma by evaluating actionable mutations and biomarkers to guide precision oncology and region-specific treatment strategies.

Study methodology

Sample identification

The cohort of 263 sarcoma patients was chosen from an initial cohort of 1749 patients who were either walk-in patients or those referred for genomic sequencing [Figure 1]. Those patients with a confirmed diagnosis of sarcoma and had a tumor tissue biopsy samples of 3-5 mm diameter were included in the study, while those patients with other cancer types were excluded from the study.

Sample testing

The 263 patients with sarcoma were screened for various germline and somatic variants including SNVs, InDels, CNVs, and gene fusions using the TARGT Indiegene™ gene panel (1212 genes) and exome sequencing on the Illumina sequencing platform. Additionally, immunotherapy biomarkers tumor mutation burden (TMB), microsatellite status, and *PD-L1* expression were tested. As reported in the KEYNOTE-158 trial, TMB-high is defined as >10 mutations/mb.[9,10] MSI was measured with MSI-sensor2, and a score of $\geq 15\%$ was considered as MSI-high.

Library preparation, sequencing, and bioinformatics data analysis

Formalin-fixed paraffin-embedded (FFPE) blocks with a minimum tumor surface area of ≥ 5 mm^2 with tumor content $\geq 10\%$ ($\sim$150 viable tumor cells per high power field (hpf in microscopy as per histological examination) were selected. An all prep FFPE DNA/RNA kit (Qiagen, Valencia, CA, US) was used for genomic DNA and total RNA extraction from the FFPE blocks. Quality control (QC)-qualified DNA/RNA samples were taken forward for library preparation, which included fragmentation, adapter ligation, amplification, and genomic DNA exon capture by overnight hybridization with exon-specific probes. The Agilent DNA Prep with an enrichment kit (catalog number 5191-6874) and Agilent RNA Exome kit (catalog number 5191-6874) were used for library preparation

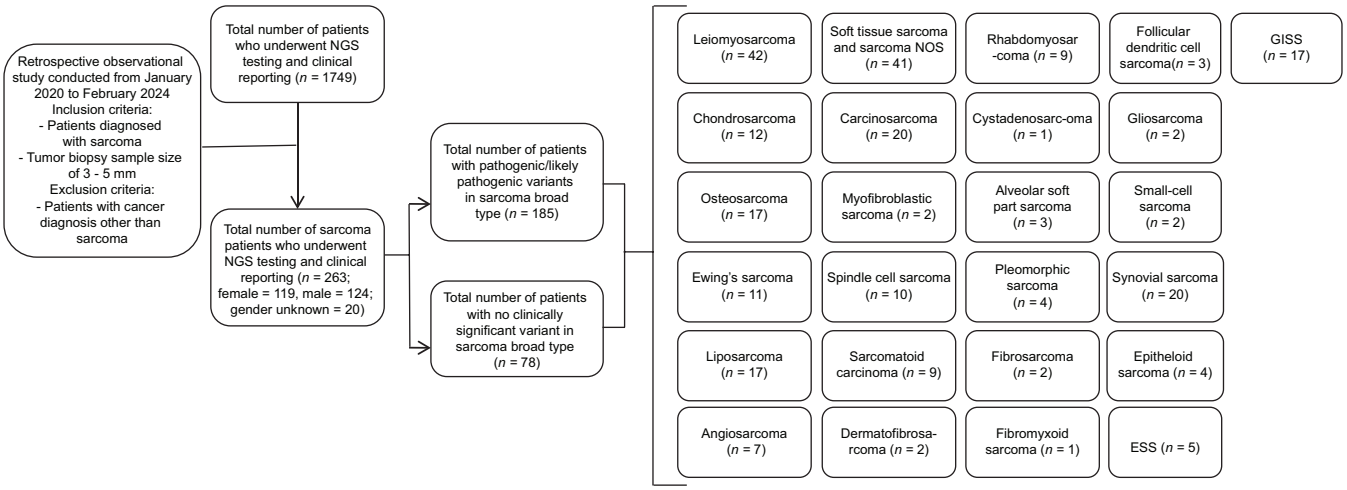


Figure 1: Flow diagram illustrating the patient selection from data repository for the study cohort. ESS = Endometrial stromal sarcoma, GISS = Gastrointestinal stromal sarcoma, NOS = Not otherwise specified

for DNA-exome and RNA-Seq (RNA-exome), respectively. The prepared libraries were QC-analyzed for fragment size and concentration; a qualified library had at least 10 nM concentration with a single distinct peak around 300 bp. The QC-qualified NGS libraries were subjected to paired-end sequencing (2 × 150 read length configuration) on NextSeq™ Systems (Illumina Inc., San Diago, CA, US) with 200X median coverage. The samples were screened using gene panels or exome sequencing.

The raw sequencing reads (FASTQ format) obtained from the high-throughput sequencer were analyzed using a customized bioinformatics pipeline to identify genomic alterations including SNVs, InDels, CNAs, and gene fusions. The Illumina DRAGEN somatic and RNA pipelines (Illumina DRAGEN Bio-IT Platform v3.6) were used for DNA and RNA exome data. The DRAGEN-aligner was used for read alignment with the hg19/GrCh37 reference sequence.

Annotated variants were filtered based on variant type, location, and its frequency in ExAC and 1000Genomes databases.[9-12] Variant annotation was done using an in-house pipeline developed with modules of population and clinical variant databases, *in silico* variant prioritization tools, complemented by a manually curated database from cBioPortal, TCGA, NCCN, FDA, CIVIC, Precision Cancer Therapy-MD Anderson, OncoKB, 7PharmGKB, clinical trials, and available literature.

Pathway analysis

The frequencies of affected pathways were assessed based on the frequencies of mutated genes belonging to the corresponding pathways. Additionally, gene set enrichment analysis (GSEA)-based pathway analysis was carried out using the molecular signatures database (MSigDB)[13] and National Cancer Institute (NCI)-Nature Pathway Interaction Database (PID).[14]

Statistics

The sample size was not calculated *a priori* as it was a retrospective study. All the obtained data were tabulated in MS Excel and presented as numbers and percentages. No statistical methods were used for analysis. The '*P* value' of statistical significance was not applicable to this study.

RESULTS

We conducted this retrospective study in a cohort of 263 patients diagnosed with sarcoma [Figure 1]. These were walk-in patients, primarily in advanced stages of disease, referred for genomic profiling following initial intervention or first/second-line therapy. These patients were chosen from an original repository of 1749 patients with various cancer types, subjected to genomic sequencing. The study cohort included 119 females (45.2%), 124 males (47.1%), and 20 patients (7.6%) whose gender was unknown; the age of patients ranged from 10 - 92 years (median 50 years). Those aged ≤30 years, categorized as pediatric, adolescent, and young adults, constituted 16.7% (44 patients) of the cohort.

Clinically significant variants

In this cohort, driver mutations or pathogenic/likely pathogenic variants (hereafter referred to as pathogenic) were detected in 185 patients (70.3%) across 138 genes [Table 1]. Pathogenic SNVs/mutations were found in 158 patients (60%), CNVs/gene amplifications were seen in 39 patients (14.8%), and gene fusions were identified in 14 patients (5.3%) [Table 2]. This included 21 germline mutations in 20 patients (7.6%). While CNVs and gene fusions were the sole driver/clinically significant variants in 21 patients (8%) and 6 patients

(2.3%), respectively, 78 patients (30%) had no pathogenic variants or driver mutations.

Among the genes harboring SNVs, 51 were tumor suppressors (43.6%) and 18 were oncogenes (15.4%), while 3 genes (2.6%) were known to play both oncogene and tumor suppressor roles. *TP53* was the most predominantly mutated gene in the present cohort, followed by *KIT, PTEN, RB1,* and *ARID1A* [Supplementary Appendix 2]. A total of 477 mutations were identified, of which 19 mutations were found to occur in ≥2 patients, several of these were variants impacting drug response such as *DPYD, SLCO1B1, MTHFR,* and *SLC19A1*, rather than sarcoma-associated.

The majority of the amplified genes were identified as oncogenes (18 genes; 78.3%). *CDK4* (6.4%; 17 patients) and *MDM2* (5.7%; 15 patients) were the most frequently amplified genes, while other genes include *PDGFRA* (1.1%; 3 patients), *FGFR1* (0.7%; 2 patients), and *DDIT3* (0.7%; 2 patients) [Table 2]. Similarly, among patients with no other driver mutations (n = 99), *CDK4* (8%; 8 patients), *MDM2* (6%; 6 patients), and *PDGFRA* (0.3%; 2 patients) were common. The other gene

Table 1: Gene mutation frequencies identified in the cohort (n=263)

Gene	Number of samples (n)	Percentage (%)	Gene	Number of samples (n)	Percentage (%)	Gene	Number of samples (n)	Percentage (%)
TP53	66	25.1	*ALDH7A1*	1	0.4	*OCA2*	1	0.4
KIT	15	5.7	*ARID2*	1	0.4	*OLFML2B*	1	0.4
MTHFR	12	4.6	*BARD1*	1	0.4	*PAH*	1	0.4
PTEN	12	4.6	*BRAF*	1	0.4	*PALB2*	1	0.4
RB1	12	4.6	*BRCA1*	1	0.4	*PDE11A*	1	0.4
ARID1A	9	3.4	*CBFB*	1	0.4	*PER1*	1	0.4
ATRX	9	3.4	*CDC73*	1	0.4	*PINK1*	1	0.4
PIK3CA	8	3	*CDKN1C*	1	0.4	*PMS2*	1	0.4
BRCA2	6	2.3	*CDKN2A*	1	0.4	*POLD1*	1	0.4
NF2	5	1.9	*CEP290*	1	0.4	*POLG*	1	0.4
TSC2	5	1.9	*CHAT*	1	0.4	*PRSS56*	1	0.4
APC	4	1.5	*CHRNG*	1	0.4	*RAD51B*	1	0.4
FBXW7	4	1.5	*CTRC*	1	0.4	*RAD54B*	1	0.4
ABCG2	3	1.1	*DIS3*	1	0.4	*RASA1*	1	0.4
BCOR	3	1.1	*DNMT3A*	1	0.4	*RET*	1	0.4
CHEK2	3	1.1	*ECHS1*	1	0.4	*ROS1*	1	0.4
CTNNB1	3	1.1	*EGFR*	1	0.4	*SLC26A4*	1	0.4
KEAP1	3	1.1	*EPHB2*	1	0.4	*SLC37A4*	1	0.4
KRAS	3	1.1	*FAM92A1*	1	0.4	*SMARCA4*	1	0.4
MSH3	3	1.1	*FANCA*	1	0.4	*SOX17*	1	0.4
NF1	3	1.1	*FANCD2*	1	0.4	*SOX2*	1	0.4
RAD50	3	1.1	*FGFR2*	1	0.4	*SPG7*	1	0.4
ATM	2	0.8	*GDF6*	1	0.4	*STAG2*	1	0.4
BCHE	2	0.8	*GJB2*	1	0.4	*SUZ12*	1	0.4
DICER1	2	0.8	*GJB4*	1	0.4	*TBXAS1*	1	0.4
ERCC2	2	0.8	*HMBS*	1	0.4	*TSHR*	1	0.4
FLCN	2	0.8	*IDH1*	1	0.4	*UGT1A1*	1	0.4
GNAS	2	0.8	*JAK1*	1	0.4	*VKORC1*	1	0.4
KMT2C	2	0.8	*KRT8*	1	0.4	*XPO1*	1	0.4
KMT2D	2	0.8	*LATS2*	1	0.4	*DPYD*	17	6.5
MEN1	2	0.8	*LIRF*	1	0.4	*SLC19A1*	14	5.3
NUBPL	2	0.8	*MAGEC3*	1	0.4	*MTHFR*	12	4.6
PADI3	2	0.8	*MAP3K1*	1	0.4	*SLCO1B1*	5	1.9
PDGFRA	2	0.8	*MLH1*	1	0.4	*CYP2D6*	4	1.5
PIK3R1	2	0.8	*MRE11A*	1	0.4	*CASP8*	1	0.4
SETD2	2	0.8	*MSH2*	1	0.4			
STK11	2	0.8	*MVK*	1	0.4			
TERT	2	0.8	*MYH7*	1	0.4			
TSC1	2	0.8	*NOTCH1*	1	0.4			
ABCC6	1	0.4	*NOTCH4*	1	0.4			
AKT1	1	0.4	*NRAS*	1	0.4			

Table 2: Gene amplifications and fusions observed in the cohort (*n*=263)

Gene amplification	Number of patients (*n*)	Percentage (%)	Gene fusion	Number of patients (*n*)	Percentage (%)
CDK4	17	6.4	*SS18-SSX2*	3	1.1
MDM2	15	5.7	*ASPSCR1-TFE3*	2	0.7
HMGA4	1	0.3	*EWSR1-FLI1*	2	0.7
DDIT3	2	0.7	*EWSR1-ERG*	1	0.3
FGFR2	1	0.3	*PAX3-FOXO1*	1	0.3
SALL4	1	0.3	*HEY1-NCOA2*	1	0.3
CRKL	1	0.3	*ZC3H7B-BCOR*	1	0.3
CUL4A	1	0.3	*COL1A1-PDGFB*	1	0.3
CCND1	1	0.3			
FGF3	1	0.3			
FGF4	1	0.3			
ETV1	1	0.3			
FGFR1	2	0.7			
KRAS	1	0.3			
YAP1	1	0.3			
MAP2K4	1	0.3			
ERBB3	1	0.3			
PDGFRA	3	1.1			
HGMA2	1	0.3			
NCOA3	1	0.3			
MAPK1	1	0.3			
GLI1	1	0.3			
CCND3	1	0.3			

amplifications observed were *ERBB3* ((0.3%; 1 patient), *FGFR1* (0.7%; 2 patients), *CCND3* (1%; 1 patient), and *CRKL* (0.3%; 1 patient). Gene fusions *SS18-SSX2* (1.1%; 3 patients), *ASPSCR1-TFE3* (0.7%; 2 patients), and *EWSR1-FLI1* (0.7%; 2 patients) were found to recur in the cohort [Table 2]. Similar frequencies were observed among patients with no drivers.

Clinically significant variants, defined as variants with FDA/NCCN approved therapy and potential targetable variants studied in clinical trials, were identified in the cohort. A total of 156 patients (59.3%) were found to have variants of therapeutic significance; this included genomic alterations such as SNVs, CNVs, fusions, and immunology biomarkers TMB-H, MSI-H, and positive PD-L1 expression. 23 patients (8.7%) had mutations that could be targeted with FDA/NCCN approved therapy. The remaining 133 patients (50.6%) had variants that are reportedly potential therapeutic targets studied by various clinical trials; of these, 10 patients (3.8%) had targets with clinical trials being assessed for the same cancer type.

Immunotherapy biomarkers

The cohort was screened for high tumor mutation burden (TMB-H), microsatellite instability (MSI-H), and positive *PD-L1* expression to understand the possibility of utilizing immunotherapy in these patients. The overall frequencies of TMB-H, MSI-H, and *PD-L1* positive patients were lower in the present cohort. Approximately, 13.3% of the cohort (35 patients) was found to be TMB-H (TMB >10), and the median TMB for the cohort was 7. MSI-H phenotype (MSI >15) was observed in 1.5% of the cohort (4 patients), while 6.8% (18 patients) were *PD-L1*-positive. TMB-H and MSI-H were found to co-occur in 0.7% of the cohort (2 patients), and 0.3% of the cohort (1 patient) had both MSI-H and *PD-L1*-positive phenotypes, while none of the patients had co-occurring TMB-H, MSI-H, and *PD-L1*-positive phenotypes.

Pathway analysis

Assessing the respective pathways of genes harboring mutations and amplifications in the cohort identified 19 different signaling pathways [Figure 2]. A majority of the cohort had mutations/amplifications in the cell cycle control pathway genes (39.2%; 103 patients), predominantly in *TP53* (25.1%; 66 patients) and *RB1* (4.6%; 12 patients). This was followed by the *PI3K/AKT/mTOR* pathway (17.9%; 47 patients) with prevalence in *KIT* (5.7%; 15 patients) and *PTEN* (4.6%; 12 patients) mutations and *RAS/RAF/MAPK* pathway genes (14.8%; 39 patients) with *ARID1A* (3.4%; 9 patients), *ATRX* (3.4%; 9 patients), and *NF2* (1.9%; 5 patients) found commonly mutated. The other commonly affected pathways were HRR pathway, DNA damage/repair pathway, chromatin remodeling, and β-catenin/WNT signaling pathways [Supplementary Appendix 3]. The cell cycle control (7.2%; 19 patients) and DNA damage/repair pathways (5.7%; 15 patients) harbored the

highest number of gene amplifications, namely, *CDK4* (6.5%; 17 patients) and *MDM2* (5.7%; 15 patients); this was followed by amplifications in the *RTK*/growth factor signaling (1.9%; 5 patients) and *RAS/RAF/MAPK* pathway (1.5%; 4 patients) genes.

Gene set enrichment analysis (GSEA) found that a majority of the mutated genes were direct p53 effectors, players in the *BARD1* signaling events, *ErbB1* downstream signaling, Fanconi anemia pathway, *PDGFRA*-β signaling pathway, apical junction, mitotic spindle, and myogenesis. Interestingly, we also found that several genes (35 genes in MSigDB, 39 genes in NCI Nature PID) were annotated for more than one pathway. *TP53*, *RAD50*, *BARD1*, and *BRCA1*, which are part of *BARD1* signaling, were also identified as E2F targets. Certain lesser-known genes, *CHRNG* and *MYH7*, which were predicted to be a part of myogenesis were also found to be part of downstream *KRAS* signaling. *PIK3CA* and *PIK3R1* were found to be a part of more than 60 signaling events, including those mediated by the hedgehog family [Supplementary Appendices 4 and 5].

Genomic variants in sarcoma subtypes

A total of 25 sarcoma subtypes were identified in our cohort. Leiomyosarcoma (LMS; 16%; 42 patients) and soft tissue sarcoma/sarcoma NOS (not otherwise specified) (15.6%; 41 patients) were the most prevalent, followed by synovial sarcoma, carcinosarcoma (7.6%, each; 20 patients), osteosarcoma, and gastrointestinal stromal sarcoma (6.5%, each; 17 patients) [Supplementary Appendix 6]. A wide range of patient ages was observed, which varied across sarcoma types [Figure 3]. Furthermore, diverse gene mutation distribution, frequencies, and mutational burden were observed in each sarcoma type [Figure 4]. *TP53* was the most predominantly mutated gene across most types of sarcomas. The frequencies of sarcoma types and cohort characteristics and corresponding genomic alterations (SNVs, CNVs, fusions) are summarized in Supplementary Appendix 7. The TMB observed across various sarcoma subtypes is represented in Supplementary Appendix 8.

Leiomyosarcomas (LMS) were the most prevalent type of sarcoma (n = 42), and the primary tumor site for 35.7% (15 patients) of these patients was the uterus (uterine leiomyosarcoma; uLMS). Mutations across 30 genes were identified in this group, where *TP53* (45.2%; 19 patients), *ATRX* (14.3%; 6 patients), *RB1* (9.5%; 4 patients), *PTEN* (7.1%; 3 patients), and *DICER1* (4.8%; 2 patients) were most prevalent. *RB1* mutations were absent among the uLMS patients. One patient (2.4%) with uterine and breast leiomyosarcoma was found to harbor a germline mutation in *MRE11A*. Co-occurring *CDK4* and *MDM2* amplifications were seen in an LMS patient (2.4%; 1 patient), *ETV1* (2.4%; 1 patient) and *MAP2K4* (2.4%; 1 patient). Out of 12 chondrosarcoma patients, only 3 (25%) were found to harbor genomic alterations, including mutations in *IDH1* (8.3%; 1 patient), *NF1* (8.3%; 1 patient), and the well-known *HEY1-NCOA2* fusion in a patient with metastatic mesenchymal chondrosarcoma (8.3%; 1 patient). Osteosarcoma patients (n = 17) were found to harbor mutations in 17 genes; *TP53* (17.6%; 3 patients) was most prevalent, while other mutations were single occurrences in genes including *RB1, AKT1,* and *CHEK2* (5.9%; 1 patient) and in unconventional genes such as *GJB2* and *GJB6*. However, there

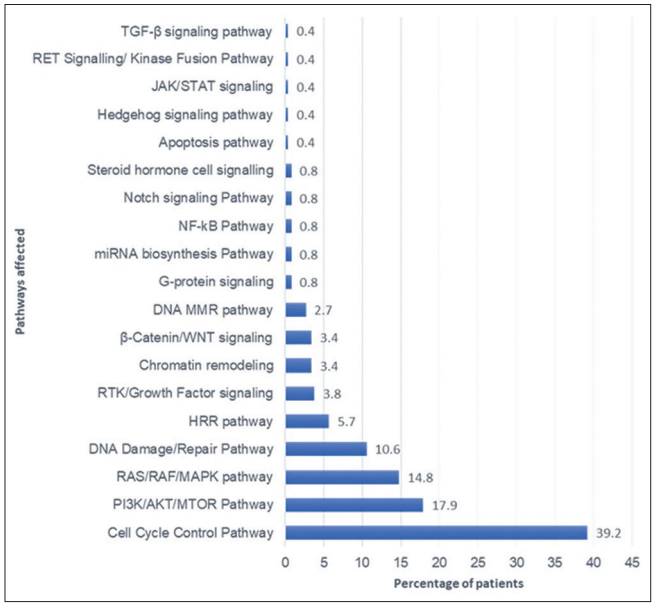


Figure 2: Pathway analysis representing the affected pathways in the study cohort

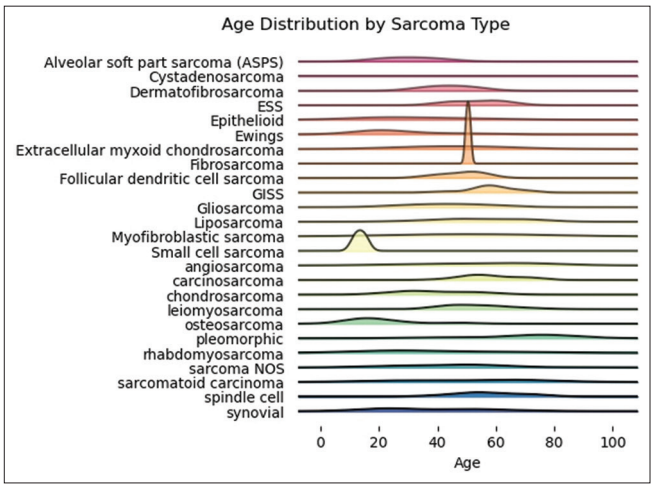


Figure 3: Density curves depicting the age distribution of patients in various sarcoma subtypes. The X-axis represents age in years, and Y-axis represents the sarcoma subtypes. The peaks for each sarcoma subtype represent the median age of patients within that specific subgroup (the median ages are given in Supplementary Appendix 7). ESS = Endometrial stromal sarcoma, GISS = Gastrointestinal stromal sarcoma

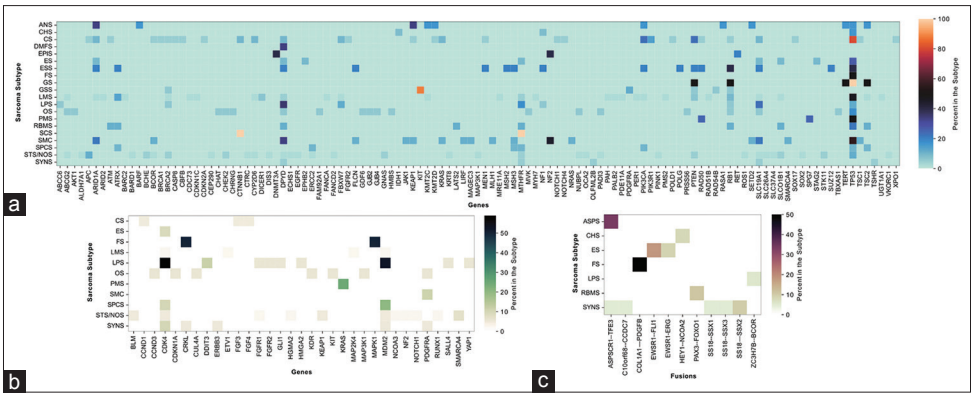


Figure 4: Heatmaps representing the various genomic alterations, including (a) SNVs/InDels, (b) Gene amplifications, and (c) Gene fusions, detected across the different sarcoma subtypes in the cohort

were several gene amplifications such as *PDGFRA, CCND3, CDKN1A, CUL4A, KIT, KDR,* and *MAP3K1* (5.9%; 1 patient). One patient (5.9%) with no other drivers had co-occurring gene amplifications in *KIT, KDR*, and *MAP3K1*. Similarly, among Ewing's sarcoma patients (n = 11), *TP53* mutations were prevalent; other mutations were seen in *ARID1A* and *STAG2* (9.1%; 1 patient). Additionally, *CDK4* amplification (9.1%; 1 patient) and typical fusions *EWSR1-ERG* (9.1%; 1 patient) and *EWSR1-FLI1* (18.2%; 2 patients) were also seen. In angiosarcoma patients (n = 7), *ARID1A* and *KEAP1* (28.6%; 2 patients) were commonly seen, while other genes include *BRAF, PIK3CA,* and *TERT* (14.3%; 1 patient).

The rhabdomyosarcoma (RMS) patients (n = 9) harbored mutations in genes including *TP53* (22.2%; 2 patients), *RB1*, *ATM,* and *ATRX* (11.1%; 1 patient). Pleomorphic RMS (11.1%; 1 patient) and alveolar RMS (22.2%; 2 patients) were the two subtypes identified in our cohort. *PAX3-FOXO1* fusion was seen in a patient (11.1%; 1 patient) with alveolar RMS. Spindle cell sarcoma patients (n = 10) had *CDK4* (10%; 1 patient) and *MDM2* (20%; 2 patients) gene amplifications, while mutations were found across genes such as *MSH3, ATRX,* and *PIK3CA* (10%; 1 patient). Pleomorphic sarcoma (n = 4) samples had mutations in *RAD50* (25%; 1 patient) and *SPG7* (25%; 1 patient) other than *TP53* (50%; 2 patients); *KRAS* gene amplification was seen in 1 patient (25%). Among the carcinosarcoma group (n = 20), 1 patient was found to have co-occurring gene amplifications in *CCND1, FGF3*, and *FGF4*, while another patient was found to harbor mutations in 12 genes (including rare genes such as *XPO1*) and had a high TMB (131.8). Mutations were seen across 29 genes with *TP53* (80%; 16 patients), *PIK3CA, PTEN* (20%, each; 4 patients), *ARIDA, CTNNB1, CYP2D6, PIK3R1,* and *KRAS* (10%, each; 2 patients) being predominant in these patients. Sarcomatoid carcinoma patients (n = 9) predominantly had mutations in *TP53* (55.5%; 5 patients), *NF2* (44.4%; 4 patients), and *ARID1A* (22.2%; 2 patients); other mutations were seen among 16 genes including *MLH1, MSH3, SETD2, BCOR*, and *KEAP1*; 1 patient (11.1%) had *PDGFRA* gene amplification.

Gene amplifications and fusions were more commonly seen than SNVs in liposarcoma, synovial sarcoma, fibrosarcoma, and alveolar soft part sarcoma (ASPS). Liposarcoma group (n = 17) could be subgrouped into dedifferentiated (41.2%; 7 patients), pleomorphic (23.5%; 4 patients), and myxoid (5.8%; 1 patient) liposarcomas. Gene amplifications were seen in 64.7% (n = 11) of all liposarcoma patients, namely, *CDK4* (58.9%; 10 patients), *MDM2* (53%; 9 patients), *DDIT3* (11.7%; 2 patients), *FGFR1, FGFR2, YAP2, GLI1,* and *HMGA2* (5.8%; 1 patient). *CDK4+MDM2* (29.4%; 5 patients) and *CDK4+DDIT3* (11.7%; 2 patients) amplifications were frequently co-occurring in these patients. The other co-amplifications include *CDK4+MDM2+DDIT3+FGF,CDK4+DDIT3+GLI, CDK4+MDM2+YAP1,* and *CDK4+MDM2+HMGA2*. *TP53* (11.7%; 2 patients) was most frequent, followed by *BRCA2, RB1, NF1,* and *FBXW7* (5.8%; 1 patient); additionally, *ZC3H7B-BCOR* fusion was seen in one patient (5.8%; 1 patient).

Among synovial sarcomas (n = 20), 40% of the patients (n = 8) had gene amplifications and fusions. Gene amplifications were seen in *CDK4* (25%; 2 patients), *MDM2* (12.5%; 1 patient), *ERBB3* (12.5%; 1 patient), *PDGFRA* (12.5%; 1 patient), and *CRKL* (12.5%; 1 patient), while fusions identified were *SS18-SSX1* (12.5%; 1 patient), *SS18-SSX2* (25%; 2 patients), *SS18-SSX3* (12.5%; 1 patient), *C10orf68-CCDC7* (12.5%; 1 patient), and *ASPSCR1-TFE3* (12.5%; 1 patient). Mutations in genes *OLFML2B* and *TSHR* (12.5%, each; 1 patient) were also found. In fibrosarcoma patients (n = 2), apart from the *TP53* mutation in one patient (50%), gene amplifications were seen in *CRKL* and *MAPK1* (50%; 1 patient); the *COL1A1-PDGFB* fusion was seen in one patient. In the alveolar soft part sarcoma (ASPS) group (n = 3), one patient had the fusion *ASPSCR1-TFE3* (33.3%).

In contrast, no gene amplifications and fusions were found in gastrointestinal stromal sarcoma (GIST), endometrial stromal

sarcoma (ESS), epithelioid sarcoma, gliosarcoma, and small cell sarcoma patients. Among GIST patients (n = 17), *KIT* (88.2%; 15 patients) and *PDGFRA* (11.8%; 2 patients) were the most predominantly mutated, occurring in 15 patients (88.2%), while other genes included *RB1* and *BRCA2* (5.9%, each; 1 patient); only 2 patients (11.7%) were negative for both *KIT* and *PDGFRA*. Although only 3 (60%) of all ESS patients (n = 5) harbored mutations, there was genomic heterogeneity with mutations seen across 14 genes. This could be due to the presence of mutations in *MMR*, *HRR*, and polymerase genes. *TP53* and *RB1* were found in 2 patients (40%), and others were single occurrences in genes including *ATRX, MSH2, MSH3, POLG, PTEN,* and *RAD50*. Epithelioid sarcoma patients (n = 4) had mutations in *DNMT3A, RET,* and *NF2* (25%, each; 1 patient). Gliosarcoma patients (n = 2) harbored mutations in *TP53* (100%), *RB1*, *TERT, PTEN*, and *TSC2* (50%, each; 1 patient). Small cell sarcoma patients (n = 2) had mutations in *CTNNB1* (50%; 1 patient). No causative genomic alterations were found in patients with follicular dendritic cell sarcoma (n = 3), dermatofibrosarcoma (n = 2), myofibroblastic sarcoma (n = 2), cystadenosarcoma (n = 1), and fibromyxoid sarcoma (n = 1).

DISCUSSION

The present study is the first to explore the genomic landscape of sarcoma using NGS-based comprehensive profiling in an Indian cohort. The study encompasses a heterogeneous cohort of 263 patients with 25 types of sarcomas, including several rare subtypes whose mutational profiles are not well characterized. We identified driver mutations and/or pathogenic variants in 70% (n = 185) of the cohort, while clinically actionable variants with therapeutic significance (FDA/NCCN approved and clinical trials) were identified in 59.3% (n = 156) of the cohort. Previous studies have reported efficacies of detecting therapeutically significant mutations ranging around 30-50% owing to the sample size and testing panel size (~400 genes).[6,7,14,15] The highest number of gene cover in the panels used in our cohort is 1212 genes, which is a significant factor in identifying targetable mutations in ~60% (n = 158) of the cohort.

It is well known that *TP53* was the most predominantly mutated gene;[6,16] similarly, in our cohort, *TP53* was the most prevalent, followed by *KIT*, *PTEN*, *RB1*, and *ARID1A*. Gene amplifications in *CDK4* and *MDM2* were prevalent in the present cohort, which is in accordance with previous studies.[6] Co-occurring *CDK4* and *MDM2* gene amplifications are associated with high-grade dedifferentiated liposarcomas;[17] however, in the present cohort, they were seen in 2.3% (n = 6) of the cohort among patients with liposarcoma, leiomyosarcoma, and synovial and spindle cell sarcomas.

It has been remarked that a 'one-size-fits-all' approach is not feasible for sarcoma and there is a need to thoroughly characterize the sarcoma subtype to identify potential drivers with therapeutic significance.[16] A 'divide-and-conquer' strategy to understand and design clinical trials to identify drivers/pathogenic targetable variants has been proposed for effective disease management in sarcoma.[17] Genomic variants specific to the different sarcoma subtypes were assessed to get insights into their pathogenic mechanisms and to identify potential therapeutic targets. Leiomyosarcomas constituted 16% of the cohort (42 patients), and uterine leiomyosarcomas were the predominant subgroup of leiomyosarcomas. Uterine leiomyosarcomas (uLMSs) are known for their aggressive nature characterized by advanced disease presentation and metastasis. Previous studies have reported mutations predominantly in *TP53, RB1, ATRX,* and *PTEN*.[18] Both LMS and uLMS had similar mutation profiles predominantly including genes *TP53, ATRX*, and *PTEN*; however, *RB1* was a prevalent mutation in LMS, while it was absent in the latter. *MRE11A* germline mutation was identified in a patient with uLMS. A chondrosarcoma patient had *IDH1* mutation in the cohort. *IDH1* mutations are common in chondrosarcomas; however, contradictory reports have emerged on their prognosis and response to various therapies, including PARPi and mTOR inhibition.[19] We noted numerous gene amplifications in the osteosarcoma group; one patient (0.4%) had co-occurring gene amplifications in RTK genes *KIT* and *KDR,* while another patient had a *PDGFRA* amplification. A pan-cancer analysis on co-amplifications in *KIT, KDR,* and *PDGFRA* (4q12amp) reported the prevalence of these amplifications in osteosarcomas and treatment with TKI monotherapy in four patients showed stable disease for >20 months.[20] Liposarcoma patients had several gene amplifications including *CDK4, MDM2, DDIT3,* and *FGFR1*. One patient with dedifferentiated liposarcoma had co-occurring amplifications in *CDK4, MDM2, DDIT3,* and *FGF*; another patient had *CDK4, DDIT3,* and *GLI1*. These occur in 12q13-15 region and are reported in mesenchymal neoplasms.[21] Although therapeutic options are being assessed for these three markers individually, the efficacy and safety of combined therapy needs to be studied. Furthermore, *YAP1* fusions have been reported previously;[22] however, *CDK4*+*MDM2*+*YAP1* co-amplifications have not been reported. The *SS18-SSX*1/2/3 gene fusions were seen in synovial sarcoma patients. *SMARCA4* amplifications are common in sarcoma and sarcoma, not otherwise specified (NOS) group. They are generally identified in ovarian cancers;[23,24] however, their prognostic and therapeutic significance is not well known. In patients with GIST, *KIT* and *PDFRA* mutations were predominant, as reported in earlier studies.[25]

Using a comprehensive panel of 1212 genes identified mutations in lesser-known genes which could potentially have therapeutic implications, such as *XPO1, MRE11A,* and *ECHS1*. *In vitro* studies have shown that the *XPO1* inhibitor selinexor

hinders tumor cell growth in dedifferentiated liposarcoma.[26] *MRE11A* was identified as a germline variant in a patient with uterine and breast leiomyosarcoma in our cohort. *MRE11A* is reportedly a negative regulator of DNA mismatch repair.[27] Pembrolizumab is known to be effective in MMR-deficient colorectal cancers (KEYNOTE-164)[28]; studying the efficacy of pembrolizumab in *MRE11A* mutated patients would be interesting. Furthermore, *MRE11A1* mutations result in HRD and therefore could be sensitive to PARPi therapy. *ECHS1* plays a role in the *PI3K/AKT/mTOR* pathway; it has been reported that targeting *ECHS1* in combination with *mTOR* inhibitors can be beneficial.[29]

Evaluating the distribution of mutated genes across pathways or cellular processes using GSEA showed that several genes played a role in more than one signaling pathway. Although conventionally used in gene expression data, GSEA identified the role played by these mutated genes in multiple pathways or cellular processes. This suggests the possibility of targeting these genes with alternative pathway-specific therapeutic agents, and conversely identifying targetable molecules in signaling cascades could be beneficial in the clinical setting. However, detailed studies are needed to validate the specificity and efficacy.

Immunotherapy with checkpoint inhibitors is reportedly well tolerated in patients with advanced sarcoma ($n = 50$), with a median overall survival of 13.4 months and a median progression-free survival of 2.4 months. [8] Therefore, the biomarkers, TMB-H, MSI-H, and *PDL1*, were screened. The median TMB is reportedly low in sarcomas;[30] a large cohort study ($n = 7494$) reported median TMB of 2.4 in MMR-proficient tumors, while for MMR-D tumors median TMB was 6.5. Similarly, the study reported lower MSI-H in 0.29%.[6] The median TMB in our cohort was 7, while MSI-H was observed in 1.5% of cases, and 6.8% were positive for *PDL1*. This could be due to the smaller sample size or the inherent nature of the study cohort. Our findings highlight the clinical value of genomic profiling in tailoring precision medicine for sarcoma treatment.

CONCLUSION

Next-generation sequencing (NGS) is beneficial in identifying clinically actionable mutations and in guiding targeted therapy in sarcoma in the form of approved guideline-based therapy, off-label therapy, or enrolment for clinical trials. It has been well established that precision medicine is the way forward owing and diagnostic testing for targetable variants facilitates improved patient outcomes by identifying appropriate therapy and reducing the unnecessary exposure to ineffective treatment by detecting druggable targets and mutations that affect therapeutic efficacy (such as presence of resistance mutations). While the cost efficacy and feasibility of using NGS for all the patients with sarcoma has been questioned due to the presence of a significant proportion of patients with no drivers identified, it seems to be the plausible choice in patients with limited therapeutic options and in advanced stages where surgery is not feasible. Identifying genes/variants specific for each sarcoma subtype and screening for those variants could be a more efficient approach for screening patients at the clinical level. However, in order to achieve that, there is a need for more studies to delineate the genomic landscape of sarcomas. Owing to its rarity and complexity, there are very few FDA/NCCN approved therapeutic options available for sarcomas. Several clinical trials are screening for efficacy of targeted therapy in solid tumors, including sarcomas; however, studies focusing specifically on therapeutic options for genomic variants in sarcoma and its subtypes are scarce. Designing preclinical studies and clinical trials to explore these aspects will aid improved outcomes in patients with sarcoma.

Acknowledgments

The authors thank the patients and clinicians for being a part of the study and providing written informed consent to use de-identified data for scientific publications.

Author contributions

Providing study samples: PSC, RV, MZ, AR, AP, KP, RV, MK, AM, AS, CBA, DS, FA, PSD, MW, VGG, BC, SR, LMS, AK, DM, AKS, NL, NL, RK, AZ, AG, AT, IS, JA, MS, MT, MS, RKD, SL, UD, VA, VP, AS, KP, VS, PJ, PC, SA, SC; study conception and design: VHV; data collection: SRP; analysis and interpretation: PVS; manuscript writing: PVS, VHV; critical revision of the article: GP, KDR, HMG; approval of the final article: all authors; accountability for all aspects of work: all authors.

Data sharing statement

Individual de-identified participant data will be made available on reasonable request, from Dr. Vidya H Veldore (vidya@4basecare.com), starting from the date of publication, until 10 years after publication. Requests beyond this timeframe will be considered on a case-by-case basis. In addition, the study protocol, including the statistical plan are already available as a supplementary appendix attached to this manuscript.

Financial support and sponsorship

The present retrospective observational study was supported by internal funding from 4basecare Precision Health Pvt. Ltd., Bengaluru, India.

Conflicts of interest

Aju Matthew and Kumar Prabhash are members of the

editorial board of Cancer Research, Statistics and Treatment. As such, they may have had access to information and/or participated in decisions that could be perceived as influencing the publication of this manuscript. However, they had recused themselves from the peer review, editorial, and decision-making process for this manuscript to ensure that the content is objective and unbiased.

REFERENCES

1. Mailankody S, Bajpai J, Budukh A, Swaminathan R, Dikshit R, Dhimal M, *et al*. Epidemiology of rare cancers in India and South Asian countries - Remembering the forgotten. Lancet Reg Health Southeast Asia 2023;12:100168.
2. WHO Classification of Tumors Editorial Board. Soft tissue and bone tumours. In: WHO Classification of Tumours. 5th edition. 2020;3. Available from: https://publications.iarc.fr/Book-And-Report-Series/Who-Classification-Of-Tumours/Soft-Tissue-And-Bone-Tumours-2020.
3. Sbaraglia M, Bellan E, Dei Tos AP. The 2020 WHO classification of soft tissue tumours: News and perspectives. Pathologica 2020;113:70-84.
4. de Juan Ferré A, Álvarez R, Casado Herráez A, Cruz Jurado J, Estival González A, Martín-Broto J, *et al*. SEOM Clinical guideline of management of soft-tissue sarcoma (2020). Clin Transl Oncol 2021;23:922-30.
5. Italiano A, Dinart D, Soubeyran I, Bellera C, Espérou H, Delmas C, *et al*. Molecular profiling of advanced soft-tissue sarcomas: The MULTISARC randomized trial. BMC Cancer 2021;21:1180.
6. Gounder MM, Agaram NP, Trabucco SE, Robinson V, Ferraro RA, Millis SZ, *et al*. Clinical genomic profiling in the management of patients with soft tissue and bone sarcoma. Nat Commun 2022;13:3406.
7. Boddu S, Walko CM, Bienasz S, Bui MM, Henderson-Jackson E, Naghavi AO, *et al*. Clinical utility of genomic profiling in the treatment of advanced sarcomas: A single-center experience. JCO Precis Oncol 2018;2:1-8.
8. Groisberg R, Hong DS, Behrang A, Hess K, Janku F, Piha-Paul S, *et al*. Characteristics and outcomes of patients with advanced sarcoma enrolled in early phase immunotherapy trials. J Immunother Cancer 2017;5:100.
9. Meléndez B, van Campenhout C, Rorive S, Remmelink M, Salmon I, D'Haene N. Methods of measurement for tumor mutational burden in tumor tissue. Transl Lung Cancer Res 2018;7:661-7.
10. Marcus L, Fashoyin-Aje LA, Donoghue M, Yuan M, Rodriguez L, Gallagher PS, *et al*. FDA approval summary: Pembrolizumab for the treatment of tumor mutational burden-high solid tumors. Clin Cancer Res 2021;27:4685-9.
11. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491:56-65.
12. Karczewski KJ, Weisburd B, Thomas B, Solomonson M, Ruderfer DM, Kavanagh D, *et al*. The ExAC browser: Displaying reference data information from over 60 000 exomes. Nucleic Acids Res 2017;45:D840-5.
13. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database hallmark gene set collection. Cell Syst 2015;1:417-25.
14. Krupa S, Anthony K, Buchoff J, Day M, Hannay T, Schaefer C. The NCI-Nature pathway interaction database: A cell signaling resource. Nat Preced 2007:1. doi: 10.1038/npre.2007.1311.1.
15. Jour G, Scarborough JD, Jones RL, Loggers E, Pollack SM, Pritchard CC, *et al*. Molecular profiling of soft tissue sarcomas using next-generation sequencing: A pilot study toward precision therapeutics. Hum Pathol 2014;45:1563-71.
16. Abu-Hijlih R, Sharaf B, Salah S, Bani Hani H, Alqaisieh M, Alzibdeh A, *et al*. Germline genetic mutations in adult patients with sarcoma: Insight into the Middle East genetic landscape. Cancers (Basel) 2024;16:1668.
17. Carmagnani PR, Groisberg R, Roszik J, Subbiah V. Precision oncology in sarcomas: Divide and conquer. JCO Precis Oncol 2019;3:PO.18.00247.
18. Kim YJ, Kim M, Park HK, Yu DB, Jung K, Song K, *et al*. Co-expression of MDM2 and CDK4 in transformed human mesenchymal stem cells causes high-grade sarcoma with a dedifferentiated liposarcoma-like morphology. Lab Invest 2019;99:1309-20.
19. Choi J, Manzano A, Dong W, Bellone S, Bonazzoli E, Zammataro L, *et al*. Integrated mutational landscape analysis of uterine leiomyosarcomas. Proc Natl Acad Sci U S A 2021;118:e2025182118.
20. Venneker S, Bovée JVMG. IDH mutations in chondrosarcoma: Case closed or not? Cancers (Basel) 2023;15:3603.
21. Disel U, Madison R, Abhishek K, Chung JH, Trabucco SE, Matos AO, *et al*. The pan-cancer landscape of coamplification of the tyrosine kinases KIT, KDR, and PDGFRA. Oncologist 2020;25:e39-47.
22. Szulzewsky F, Holland EC, Vasioukhin V. YAP1 and its fusion proteins in cancer initiation, progression and therapeutic resistance. Dev Biol 2021;475:205-21.
23. Lavernia J, Claramunt R, Romero I, López-Guerrero JA, Llombart-Bosch A, Machado I. Soft tissue sarcomas with chromosomal alterations in the 12q13-15 region: Differential diagnosis and therapeutic implications. Cancers (Basel) 2024;16:432.
24. Peng L, Li J, Wu J, Xu B, Wang Z, Giamas G, *et al*. A pan-cancer analysis of SMARCA4 alterations in human cancers. Front Immunol 2021;12:762598.
25. Joensuu H. KIT and PDGFRA Variants and the survival of patients with gastrointestinal stromal tumor treated with adjuvant imatinib. Cancers (Basel) 2023;15:3879.
26. Azmi AS, Uddin MH, Mohammad RM. The nuclear export protein XPO1 - from biology to targeted therapy. Nat Rev Clin Oncol 2021;18:152-69.
27. Du D, Yang Y, Zhang Y, Wang G, Chen L, Guan X, *et al*. MRE11A: A novel negative regulator of human DNA mismatch repair. Cell Mol Biol Lett 2024;29:37.
28. Le DT, Kim TW, Van Cutsem E, Geva R, Jäger D, Hara H, *et al*. Phase II open-label study of pembrolizumab in treatment-refractory, microsatellite instability–high/mismatch repair–deficient metastatic colorectal cancer: KEYNOTE-164. J Clin Oncol 2020;38:11-9.
29. Hu T, Chen X, Lu S, Zeng H, Guo L, Han Y. Biological role and mechanism of lipid metabolism reprogramming related gene ECHS1 in cancer. Technol Cancer Res Treat 2022;21:153303382211406.
30. Yiong CS, Lin TP, Lim VY, Toh TB, Yang VS. Biomarkers for immune checkpoint inhibition in sarcomas - are we close to clinical implementation? Biomark Res 2023;11:75.

SUPPLEMENTARY APPENDIX 1: STUDY PROTOCOL

- For this study, the blood and tissue samples will be collected from cancerous patients with informed consent, questionnaire, and clinical medical report after obtaining ethical clearance.
- Genomic DNA will be isolated from the patients' blood and Formalin-Fixed Paraffin-Embedded (FFPE) blocks.
- Quality Control (QC) qualified DNA samples were processed for library preparation, which includes fragmentation, adapter addition, amplification, and capturing of exonic regions through overnight hybridization of exon-specific probes.
- The prepared libraries underwent QC analysis for the detection of library fragment size, and concentration.
- The qualified NGS libraries were subjected to paired end (2 × 150 read length configuration) sequencing on the NextSeq™ Systems (Illumina Inc., San Diago, CA) at a mean coverage depth of 200X.

Patient recruitment/study cohort:
- A sample size of 3 mm to 5 mm of tumor or normal tissue will be taken from patients who undergo surgery or tumor biopsy (removal of a small piece of tumor) for medical reasons or as part of a research treatment protocol.
- Participants will have 5 milliliters of blood drawn at the beginning of the study.

Study Type: Observational.

Study design: Case Control.

Primary outcome measures
- Genetic analysis of tissue and blood samples for mutations.

Estimated enrolment

Walk-in cancer patients - 1749 samples.

Duration of study - 3 years.

Eligibility:

Ages eligible for study: All ages.

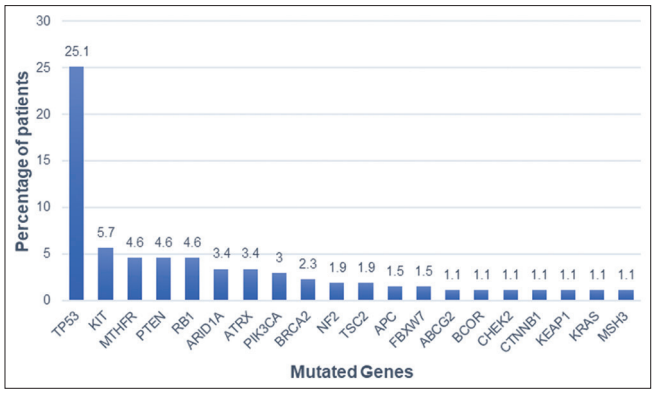
Sexes eligible for study: All (male, female).

Criteria

Inclusion criteria
1. Patients of all ages are eligible.
   The study does not provide any incentives or reimbursements to the study participants enrolled. Also, the PI does not receive any monetary benefit from any other agencies or institutions.

Exclusion criteria
1. Pregnant individuals will not be eligible due to potential risks to the fetus associated with radiologic procedures required for biopsy.

Clinical trials-not applicable.



SUPPLEMENTARY APPENDIX 2: Frequencies of the 20 predominantly mutated genes in the cohort

SUPPLEMENTARY APPENDIX 3: Frequencies of genes belonging to specific signaling pathways in the cohort (*n*=263)

Pathway	Number of patients (*n*)	Percentage (%)
Cell cycle control pathway	103	39.2
PI3K/AKT/MTOR pathway	47	17.9
RAS/RAF/MAPK pathway	39	14.8
DNA damage/repair pathway	28	10.6
HRR pathway	15	5.7
RTK/growth factor signaling	10	3.8
Chromatin remodeling	9	3.4
β-Catenin/WNT signaling	9	3.4
DNA MMR pathway	7	2.7
G-protein signaling	2	0.8
miRNA biosynthesis pathway	2	0.8
NF-kB pathway	2	0.8
Notch signaling pathway	2	0.8
Steroid hormone cell signaling	2	0.8
Apoptosis pathway	1	0.4
Hedgehog signaling pathway	1	0.4
JAK/STAT signaling	1	0.4
RET signaling/kinase fusion pathway	1	0.4
TGF-β signaling pathway	1	0.4

PI3K/AKT/MTOR=Phosphatidylinositol-3-kinase/protein kinase B/mammalian target of rapamycin pathway, RAS/RAF/MAPK=Rat sarcoma virus, rapidly accelerated fibrosarcoma, and mitogen-activated protein kinase pathway, HRR=Homologous recombination repair, RTK=Receptor tyrosine kinase; WNT=Wingless-related integration site, MMR=Mismatch repair; NF-kB=Nuclear factor-kappa B, JAK/STAT=Janus kinase-signal transducer and activation of transcription, RET=Rearranged during transfection, TGF-β = Transforming growth factor beta

SUPPLEMENTARY APPENDIX 4: Gene set enrichment analysis (GSEA) based pathway analysis

Gene	MSigDB prediction	NCI Nature PID prediction
TP53	E2F Targets | Wnt-beta Catenin signaling | p53 pathway | DNA repair	*BARD1* signaling events | Direct p53 effectors | *LKB1* signaling events | p53 pathway | *PLK3* signaling events | Aurora A signaling | AP-1 transcription factor network | p75 (NTR)-mediated signaling | Validated targets of C-MYC transcriptional activation | Hypoxic and oxygen homeostasis regulation of HIF-1-alpha | Signaling events mediated by HDAC Class III | Signaling mediated by p38-alpha and p38-beta | Glucocorticoid receptor regulatory network
DPYD	Apoptosis	NA
KIT	UV Response Dn	Signaling events mediated by stem cell factor receptor (c-Kit) | C-MYB transcription factor network
SLC19A1	Adipogenesis | Myc Targets V2	NA
MTHFR	KRAS Signaling Dn	NA
PTEN	PI3K/AKT/mTOR Signaling | Apical Junction | UV Response Dn	*PDGFR*-beta signaling pathway | Direct p53 effectors | Class I PI3K signaling events | Signaling events mediated by Stem cell factor receptor (c-Kit) | *BCR* signaling pathway | TCR signaling in naive CD4+T cells | AP-1 transcription factor network | CXCR4-mediated signaling events | RhoA signaling pathway
RB1	p53 Pathway | Myogenesis	Direct p53 effectors | p73 transcription factor network | FOXM1 transcription factor network | ATF-2 transcription factor network | E2F transcription factor network | Notch-mediated HES/HEY network | Regulation of retinoblastoma protein
ATRX	UV response Dn | G2-M checkpoint	NA
PIK3CA	Complement	ErbB2/ErbB3 signaling events | *PDGFR*-beta signaling pathway | IL2-mediated signaling events | *SHP2* signaling | *EGF* receptor (*ErbB1*) signaling pathway Homo sapiens NULL | CDC42 signaling events | Signaling events mediated by Hepatocyte Growth Factor Receptor (c-Met) | *ErbB1* downstream signaling | GMCSF-mediated signaling events | Internalization of *ErbB1* | a6b1 and a6b4 Integrin signaling | E-cadherin signaling in keratinocytes | Signaling events mediated by focal adhesion kinase | Fc-epsilon receptor I signaling in mast cells | Neurotrophic factor-mediated Trk receptor signaling | Signaling events mediated by *VEGFR1* and *VEGFR2* | N-cadherin signaling events | *EPHB* forward signaling | *CXCR3*-mediated signaling events | IL6-mediated signaling events | Class I *PI3K* signaling events | Signaling events mediated by Stem cell factor receptor (c-Kit) | BCR signaling pathway | *VEGFR1* specific signals | *IGF1* pathway | Nectin adhesion pathway | IL2 signaling events mediated by *STAT5* | *IL2* signaling events mediated by *PI3K* | Trk receptor signaling mediated by *PI3K* and PLC-gamma | *ErbB4* signaling events | Signaling events regulated by Ret tyrosine kinase | E-cadherin signaling in the nascent adherens junction | Plasma membrane estrogen receptor signaling | IFN-gamma pathway | Signaling events mediated by TCPTP | Insulin Pathway | Angiopoietin receptor Tie2-mediated signaling | IL5-mediated signaling events | *FGF* signaling pathway | Integrins in angiogenesis | Signaling events mediated by the Hedgehog family | *PDGFR*-alpha signaling pathway | *VEGFR3* signaling in lymphatic endothelium | *IL3*-mediated signaling events | TRAIL signaling pathway | Nongenotropic Androgen signaling | Ephrin B reverse signaling | *CXCR4*-mediated signaling events | Nephrin/Neph 1 signaling in the kidney podocyte | Signaling events mediated by PTP1B | IL4-mediated signaling events | p75 (NTR)-mediated signaling | Atypical NF-kappaB pathway | *EPHA2* forward signaling | Reelin signaling pathway | Netrin-mediated signaling events | Osteopontin-mediated events | IL1-mediated signaling events | IL23-mediated signaling events | FAS (CD95) signaling pathway | PAR1-mediated thrombin signaling events
BRCA2	E2F targets | mitotic spindle | G2-M checkpoint	Fanconi anemia pathway | p73 transcription factor network | *FOXM1* transcription factor network | Validated transcriptional targets of deltaNp63 isoforms | ATR signaling pathway
NF2	Apical junction | spermatogenesis	*ErbB2/ErbB3* signaling events
TSC2	PI3K/AKT/mTOR signaling | myogenesis	Direct p53 effectors | Validated targets of C-MYC transcriptional repression | mTOR signaling pathway | *LKB1* signaling events | p38 signaling mediated by *MAPKAP* kinases
APC	TGF-beta signaling | mitotic spindle	CDC42 signaling events | Signaling events mediated by Hepatocyte Growth Factor Receptor (c-Met) | Direct p53 effectors | Presenilin action in Notch and Wnt signaling | Degradation of beta catenin | Canonical Wnt signaling pathway | Regulation of nuclear beta catenin signaling and target gene transcription | Regulation of CDC42 activity
FBXW7	p53 pathway	C-MYC pathway | Notch signaling pathway
ABCG2	Heme metabolism	HIF-1-alpha transcription factor network | HIF-2-alpha transcription factor network
BCOR	NA	Signaling events mediated by HDAC Class II
CHEK2	E2F targets	ATM pathway | p53 pathway | *PLK3* signaling events | *FOXM1* transcription factor network

Contd...

SUPPLEMENTARY APPENDIX 4: Contd...

Gene	MSigDB prediction	NCI Nature PID prediction
CTNNB1	Wnt-beta catenin signaling | TGF-beta Signaling | apoptosis | cholesterol homeostasis	CDC42 signaling events | Signaling events mediated by hepatocyte growth factor receptor (c-Met) | E-cadherin signaling in keratinocytes | Signaling events mediated by *VEGFR1* and *VEGFR2* | N-cadherin signaling events | Presenilin action in Notch and Wnt signaling | Degradation of beta catenin | TGF-beta receptor signaling | Canonical Wnt signaling pathway | Nectin adhesion pathway | Regulation of nuclear beta catenin signaling and target gene transcription | E-cadherin signaling in the nascent adherens junction | Stabilization and expansion of the E-cadherin adherens junction | Posttranslational regulation of adherens junction stability and dissassembly | *RAC1* signaling pathway | AP-1 transcription factor network | *Arf6* trafficking events | Coregulation of androgen receptor activity | Integrin-linked kinase signaling | FoxO family signaling
KRAS	NA	*ErbB2/ErbB3* signaling events | *PDGFR*-beta signaling pathway | IL2-mediated signaling events | *SHP2* signaling | *EGF* receptor (*ErbB1*) signaling pathway homo sapiens NULL | *ErbB1* downstream signaling | Ras signaling in the CD4+TCR pathway | GMCSF-mediated signaling events | Internalization of *ErbB1* | Neurotrophic factor-mediated Trk receptor signaling | Regulation of Ras family activation | mTOR signaling pathway | Trk receptor signaling mediated by the *MAPK* pathway | EPHB forward signaling | CXCR3-mediated signaling events | Class I *PI3K* signaling events | TCR signaling in naive CD4+T cells | Downstream signaling in naive CD8+T cells | Trk receptor signaling mediated by *PI3K* and PLC-gamma | Plasma membrane estrogen receptor signaling | TCR signaling in naive CD8+T cells | C-MYB transcription factor network
NF1	Hedgehog signaling | apical junction | mitotic spindle	Regulation of Ras family activation | ATF-2 transcription factor network | Syndecan-2-mediated signaling events | FOXA2 and FOXA3 transcription factor networks
RAD50	E2F targets	Fanconi anemia pathway | BARD1 signaling events | ATM pathway | Regulation of telomerase
ATM	NA	Fanconi anemia pathway | BARD1 signaling events | ATM pathway | p53 pathway | Validated transcriptional targets of deltaNp63 isoforms | Regulation of telomerase | E2F transcription factor network | Canonical NF-kappaB pathway | p38 *MAPK* signaling pathway
DICER1	NA	Validated transcriptional targets of TAp63 isoforms
ERCC2	DNA repair | reactive oxygen species pathway	NA
GNAS	Protein secretion	NA
KMT2D	KRAS signaling Dn	NA
PDGFRA	NA	*PDGFR*-alpha signaling pathway | ATF-2 transcription factor network | *PDGF* receptor signaling network
PIK3R1	NA	EGF receptor (*ErbB1*) signaling pathway homo sapiens NULL | EPHB forward signaling | E-cadherin signaling in keratinocytes | *ErbB1* downstream signaling | *ErbB2/ErbB3* signaling events | IL2-mediated signaling events | SHP2 signaling | IL2 signaling events mediated by *PI3K* | BCR signaling pathway | Trk receptor signaling mediated by *PI3K* and PLC-gamma | FAS (CD95) signaling pathway | CDC42 signaling events | Plasma membrane estrogen receptor signaling | IFN-gamma pathway | Internalization of *ErbB1* | *CXCR3*-mediated signaling events | Signaling events mediated by hepatocyte growth factor receptor (c-Met) | *PDGFR*-alpha signaling pathway | Class I *PI3K* signaling events | *PDGFR*-beta signaling pathway | Signaling events mediated by stem cell factor receptor (c-Kit) | *VEGFR1* specific signals | TRAIL signaling pathway | Fc-epsilon receptor I signaling in mast cells | Neurotrophic factor-mediated Trk receptor signaling | Signaling events mediated by *VEGFR1* and *VEGFR2* | GMCSF-mediated signaling events | Signaling events regulated by Ret tyrosine kinase | E-cadherin signaling in the nascent adherens junction | Signaling events mediated by TCPTP | Insulin pathway | a6b1 and a6b4 Integrin signaling | IL6-mediated signaling events | Angiopoietin receptor Tie2-mediated signaling | Signaling events mediated by *PTP1B* | Signaling events mediated by the Hedgehog family | Signaling events mediated by focal adhesion kinase | FGF signaling pathway | *VEGFR3* signaling in lymphatic endothelium | IL4-mediated signaling events | IGF1 pathway | Nectin adhesion pathway | Reelin signaling pathway | p75(NTR)-mediated signaling | IL2 signaling events mediated by STAT5 | Nongenotropic androgen signaling | Ephrin B reverse signaling | Osteopontin-mediated events | Nephrin/Neph 1 signaling in the kidney podocyte | Integrins in angiogenesis | N-cadherin signaling events | *CXCR4*-mediated signaling events | IL5-mediated signaling events | Atypical NF-kappaB pathway | *EPHA2* forward signaling | LPA receptor mediated events | IL3-mediated signaling events | Netrin-mediated signaling events | IL1-mediated signaling events | IL23-mediated signaling events | *ErbB4* signaling events | PAR1-mediated thrombin signaling events | EPO signaling pathway
STK11	NA	*LKB1* signaling events
TERT	NA	*IL2* signaling events mediated by *PI3K* | Regulation of telomerase | Regulation of nuclear beta catenin signaling and target gene transcription | HIF-1-alpha transcription factor network | Validated targets of C-MYC transcriptional activation
TSC1	Apical junction | Mitotic spindle	mTOR signaling pathway | *LKB1* signaling events

Contd...

SUPPLEMENTARY APPENDIX 4: Contd...

Gene	MSigDB prediction	NCI Nature PID prediction
AKT1	PI3K/AKT/mTOR signaling | Allograft rejection | Androgen response	E-cadherin signaling in keratinocytes | *ErbB1* downstream signaling | *ErbB2/ErbB3* signaling events | IL2 signaling events mediated by *PI3K* | *BCR* signaling pathway | Trk receptor signaling mediated by PI3K and PLC-gamma | Regulation of telomerase | FAS (CD95) signaling pathway | mTOR signaling pathway | Plasma membrane estrogen receptor signaling | IFN-gamma pathway | CXCR3-mediated signaling events | Signaling events mediated by hepatocyte growth factor receptor (c-Met) | Signaling events mediated by stem cell factor receptor (c-Kit) | VEGFR1 specific signals | Fc-epsilon receptor I signaling in mast cells | Coregulation of androgen receptor activity | Aurora A signaling | Signaling events mediated by *VEGFR1* and *VEGFR2* | E-cadherin signaling in the nascent adherens junction | Insulin pathway | a6b1 and a6b4 Integrin signaling | IL6-mediated signaling events | Angiopoietin receptor Tie2-mediated signaling | Signaling events mediated by *PTP1B* | Signaling events mediated by the Hedgehog family | p53 pathway | FGF signaling pathway | *VEGFR3* signaling in lymphatic endothelium | IL4-mediated signaling events | TCR signaling in naive CD4+T cells | IGF1 pathway | Reelin signaling pathway | p75(NTR)-mediated signaling | Nongenotropic Androgen signaling | Nephrin/Neph 1 signaling in the kidney podocyte | Integrins in angiogenesis | Ceramide signaling pathway | Integrin-linked kinase signaling | FoxO family signaling | *CXCR4*-mediated signaling events | TCR signaling in naive CD8+T cells | Caspase cascade in apoptosis | LPA receptor mediated events | HIF-1-alpha transcription factor network | Glucocorticoid receptor regulatory network | CD40/CD40L signaling | FOXA2 and FOXA3 transcription factor networks | Hedgehog signaling events mediated by Gli proteins | Thromboxane A2 receptor signaling | Regulation of nuclear SMAD2/3 signaling | Insulin-mediated glucose transport | IL8- and CXCR1-mediated signaling events | Retinoic acid receptors-mediated signaling | S1P3 pathway | IL8- and CXCR2-mediated signaling events | Class I PI3K signaling events mediated by Akt | amb2 integrin signaling | Endothelins
ALDH7A1	Glycolysis	NA
BARD1	E2F Targets | KRAS signaling Dn | G2-M checkpoint	BARD1 signaling events
BRAF	Spermatogenesis	PDGFR-beta signaling pathway | CDC42 signaling events | *ErbB1* downstream signaling | Ras signaling in the CD4+TCR pathway | Signaling events mediated by focal adhesion kinase | Signaling events mediated by VEGFR1 and VEGFR2 | mTOR signaling pathway | Trk receptor signaling mediated by the MAPK pathway | Downstream signaling in naive CD8+T cells
BRCA1	E2F targets | Apoptosis | Allograft Rejection | Apical surface	Fanconi anemia pathway | BARD1 signaling events | Validated targets of C-MYC transcriptional repression | ATM pathway | Aurora A signaling | ATF-2 transcription factor network | E2F transcription factor network | Coregulation of androgen receptor activity | Validated nuclear estrogen receptor alpha network | FOXA1 transcription factor network
CASP8	Apoptosis | Interferon gamma response | Interferon alpha response	FAS (CD95) signaling pathway | TRAIL signaling pathway | Coregulation of androgen receptor activity | Integrins in angiogenesis | Ceramide signaling pathway | Caspase cascade in apoptosis | TNF receptor signaling pathway | HIV-1 Nef (negative effector of Fas and TNF-alpha)
CBFB	NA	ATF-2 transcription factor network | AP-1 transcription factor network | Regulation of nuclear SMAD2/3 signaling
CDKN1C	TGF-beta signaling | Hypoxia | UV response Up | IL-2/STAT5 signaling	NA
CDKN2A	E2F Targets | p53 pathway | Allograft rejection	NA
CHRNG	Myogenesis | KRAS signaling Dn	NA
CTRC	NA	Urokinase-type plasminogen activator (uPA) and uPAR-mediated signaling
DNMT3A	NA	Validated targets of C-MYC transcriptional repression
ECHS1	Adipogenesis | Fatty acid metabolism | Oxidative phosphorylation	NA
EGFR	PI3K/AKT/mTOR Signaling | Apical junction | Allograft rejection | Protein secretion | Hypoxia | Glycolysis	*SHP2* signaling | *EGF* receptor (*ErbB1*) signaling pathway homo sapiens NULL | Direct p53 effectors | *ErbB1* downstream signaling | Internalization of *ErbB1* | a6b1 and a6b4 Integrin signaling | E-cadherin signaling in keratinocytes | ErbB receptor signaling network | Stabilization and expansion of the E-cadherin adherens junction | Signaling events mediated by *TCPTP* | Post translational regulation of adherens junction stability and dissassembly | Regulation of telomerase | Signaling events mediated by *PTP1B* | *EGFR*-dependent endothelin signaling events | Syndecan-3-mediated signaling events | Arf6 signaling events | Urokinase-type plasminogen activator (uPA) and uPAR-mediated signaling | Thromboxane A2 receptor signaling | LPA receptor mediated events
EPHB2	KRAS signaling up	*EPHB* forward signaling | Ephrin B reverse signaling | Syndecan-2-mediated signaling events | EphrinB-EPHB pathway homo sapiens NULL
FANCA	NA	Fanconi anemia pathway | *BARD1* signaling events

Contd...

SUPPLEMENTARY APPENDIX 4: Contd...

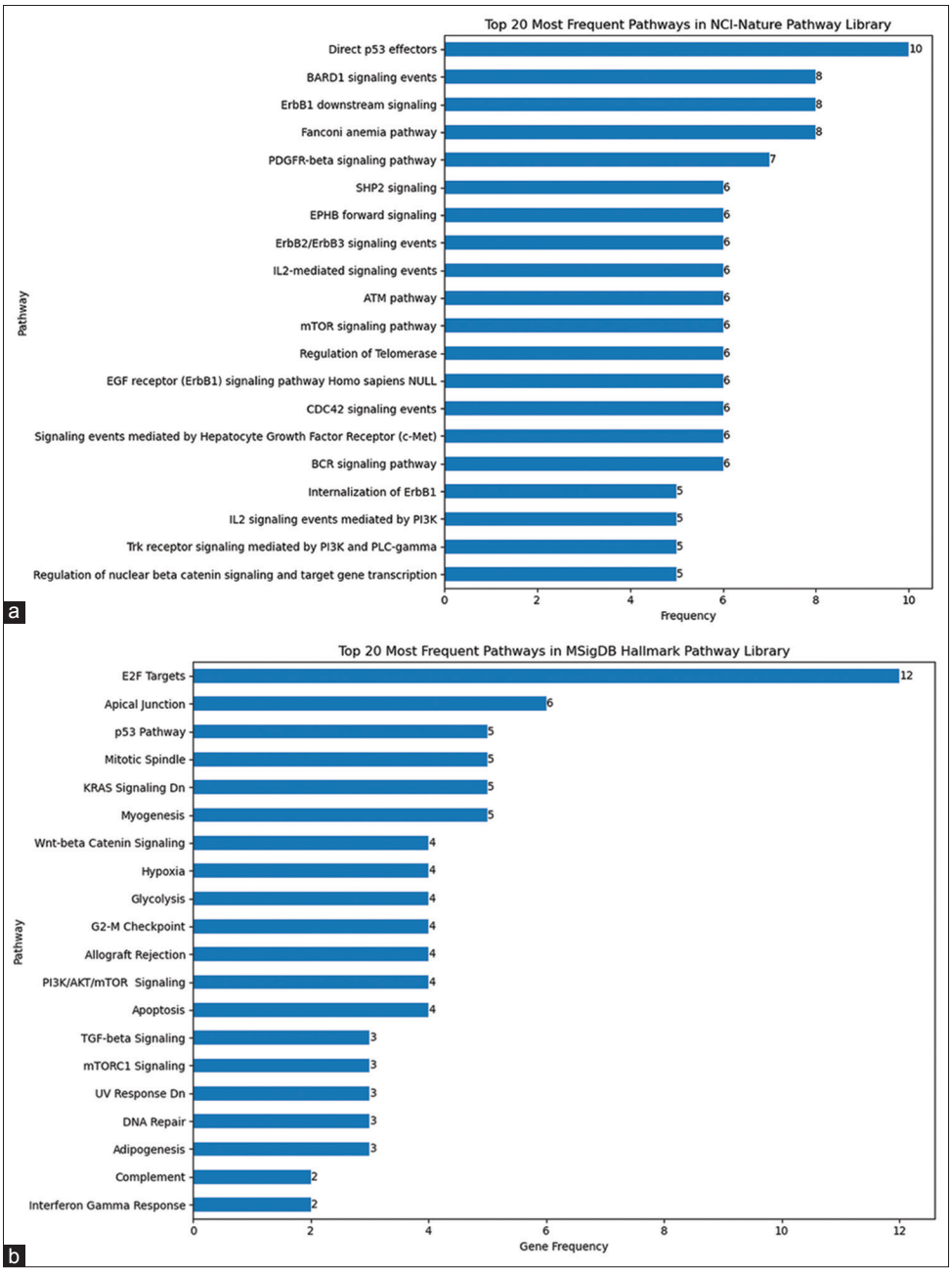
Gene	MSigDB prediction	NCI Nature PID prediction
FANCD2	NA	*BARD1* signaling events | Fanconi anemia pathway | ATM pathway | ATR signaling pathway
FGFR2	NA	FGF signaling pathway
HMBS	mTORC1 signaling | heme metabolism	NA
IDH1	Glycolysis | Adipogenesis | mTORC1 signaling | Peroxisome | Fatty acid metabolism | Xenobiotic metabolism | Oxidative phosphorylation | Bile acid metabolism	NA
JAK1	Estrogen response late	IL2-mediated signaling events | *SHP2* signaling | p73 transcription factor network | IL6-mediated signaling events | IL2 signaling events mediated by STAT5 | IL2 signaling events mediated by *PI3K* | IFN-gamma pathway | Signaling events mediated by *TCPTP* | *PDGFR*-alpha signaling pathway | IL4-mediated signaling events | IL27-mediated signaling events
KRT8	Androgen response | Estrogen response early	Signaling mediated by p38-alpha and p38-beta
LATS2	Interferon gamma response	Coregulation of androgen receptor activity
MAP3K1	Hypoxia | KRAS signaling up	*ErbB1* downstream signaling | BCR signaling pathway | FAS (CD95) signaling pathway | CDC42 signaling events | IFN-gamma pathway | Signaling events mediated by hepatocyte growth factor receptor (c-Met) | TRAIL signaling pathway | Fc-epsilon receptor I signaling in mast cells | Osteopontin-mediated events | Ceramide signaling pathway | Caspase cascade in apoptosis | p38 MAPK signaling pathway | CD40/CD40L signaling | TNF receptor signaling pathway | Role of calcineurin-dependent NFAT signaling in lymphocytes | RAC1 signaling pathway | JNK signaling in the CD4+TCR pathway | Regulation of cytoplasmic and nuclear SMAD2/3 signaling
MLH1	E2F targets	Direct p53 effectors
MRE11A	NA	*BARD1* signaling events | Fanconi anemia pathway | ATM pathway | Regulation of telomerase | Validated transcriptional targets of deltaNp63 isoforms
MSH2	E2F Targets | Peroxisome	Direct p53 effectors
MVK	Cholesterol homeostasis	NA
MYH7	Myogenesis | KRAS signaling Dn	NA
NOTCH1	Wnt-beta Catenin Signaling | p53 Pathway | Myogenesis | Notch Signaling	Presenilin action in Notch and Wnt signaling | Validated transcriptional targets of deltaNp63 isoforms | Notch-mediated HES/HEY network | Notch signaling pathway
NOTCH4	Wnt-beta Catenin Signaling | Complement	Notch signaling pathway
NRAS	NA	*ErbB2/ErbB3* signaling events | *PDGFR*-beta signaling pathway | IL2-mediated signaling events | *SHP2* signaling | *EGF* receptor (*ErbB1*) signaling pathway Homo sapiens NULL | *ErbB1* downstream signaling | Ras signaling in the CD4+TCR pathway | GMCSF-mediated signaling events | Internalization of *ErbB1* | Neurotrophic factor-mediated Trk receptor signaling | Regulation of Ras family activation | mTOR signaling pathway | Trk receptor signaling mediated by the MAPK pathway | EPHB forward signaling | CXCR3-mediated signaling events | Class I PI3K signaling events | TCR signaling in naive CD4+T cells | Downstream signaling in naive CD8+T cells | Trk receptor signaling mediated by PI3K and PLC-gamma | Plasma membrane estrogen receptor signaling | TCR signaling in naive CD8+T cells | C-MYB transcription factor network
PALB2	NA	Fanconi anemia pathway
PER1	TNF-alpha signaling via NF-kB	Circadian rhythm pathway
PINK1	Xenobiotic metabolism	NA
PMS2	E2F targets	Direct p53 effectors
POLD1	E2F targets | DNA repair	NA
RASA1	Hedgehog signaling | Apical junction | Mitotic spindle	PDGFR-beta signaling pathway | IL2-mediated signaling events | EGF receptor (*ErbB1*) signaling pathway Homo sapiens NULL | Signaling events mediated by focal adhesion kinase | Fc-epsilon receptor I signaling in mast cells | Neurotrophic factor-mediated Trk receptor signaling | Regulation of Ras family activation | EPHB forward signaling | BCR signaling pathway | *VEGFR1* specific signals | Aurora A signaling | Signaling events regulated by Ret tyrosine kinase | Insulin pathway | Angiopoietin receptor Tie2-mediated signaling | Syndecan-2-mediated signaling events | Aurora B signaling
RET	UV response up | Estrogen response early | Estrogen response late	Signaling events regulated by Ret tyrosine kinase | Post translational regulation of adherens junction stability and dissassembly
ROS1	Inflammatory response	NA
SLC37A4	Hypoxia | Glycolysis | mTORC1 signaling	NA
SMARCA4	NA	Direct p53 effectors | Regulation of nuclear beta catenin signaling and target gene transcription | Glucocorticoid receptor regulatory network | Regulation of retinoblastoma protein | Validated nuclear estrogen receptor beta network
STAG2	NA	PLK1 signaling events

Contd...

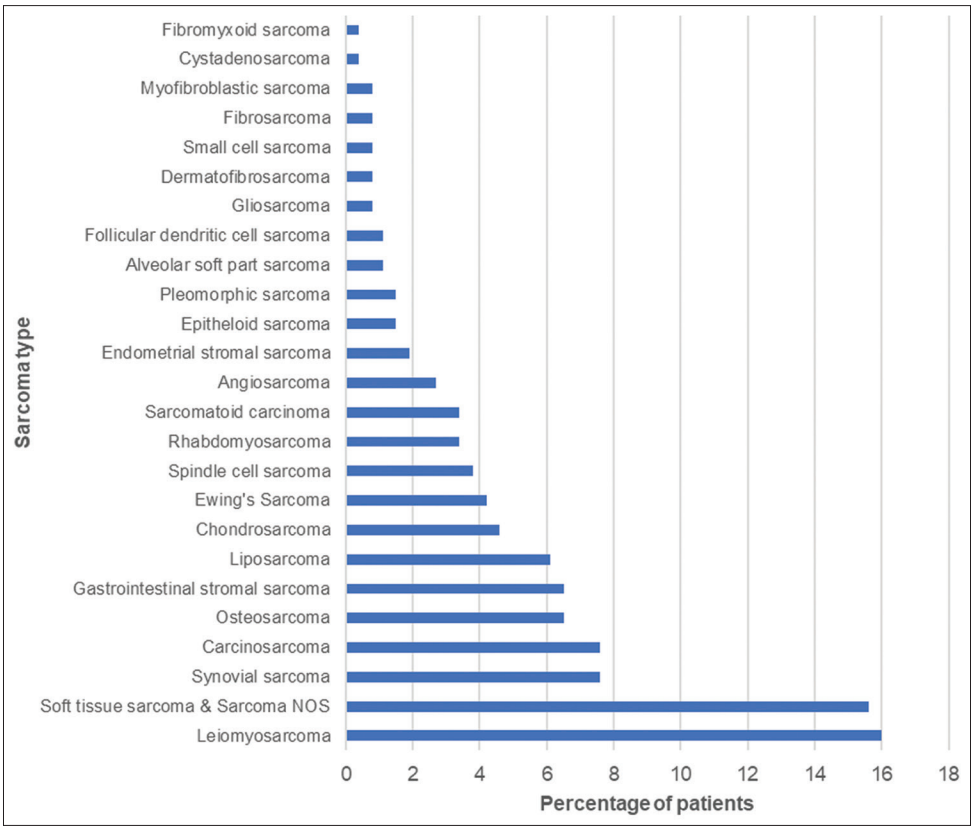
SUPPLEMENTARY APPENDIX 4: Contd...

Gene	MSigDB prediction	NCI Nature PID prediction
TSHR	NA	Arf6 signaling events|Arf6 trafficking events
XPO1	E2F targets|G2-M Checkpoint|Myc targets V1	Regulation of nuclear beta catenin signaling and target gene transcription|Integrin-linked kinase signaling|FoxO family signaling|Canonical NF-kappaB pathway|Signaling events mediated by HDAC Class II|Hedgehog signaling events mediated by Gli proteins|Role of calcineurin-dependent NFAT signaling in lymphocytes|Sumoylation by RanBP2 regulates transcriptional repression|Signaling events mediated by HDAC class I

NA=Not available



SUPPLEMENTARY APPENDIX 5: Frequencies of the 20 predominantly affected pathways identified by GSEA analysis using (a) NCI nature pathway library and (b) MSigDB pathway library



SUPPLEMENTARY APPENDIX 6: Frequency distribution of various sarcoma subtypes in the study cohort

SUPPLEMENTARY APPENDIX 7: Cohort characteristics and genomic alterations identified across sarcoma subtypes in the study cohort (*n*=263)

Sarcoma type	Number of patients	Age range; median	Gender	Subtypes	Primary sites	Metastatic sites	TMB-H, MSI-H, PD-L1 +	Genomic alterations	Frequency (%)
Leiomyosarcoma	42	20-80; 52	F=30; M=12	Uterine leiomyosarcoma (35.7%)	Uterine (35.8%), retroperitoneum and peritoneum (14.3%), ovary, thigh, seminal vesicle, cervix, kidney, para-testicular region, iliac bone, ureter (2.4% each), NOS (28.6%)	Lung, skeletal, scalp, hilar mass (2.4% each)	TMB-H=3; MSI-H=0; PDL1=3	*CDK4* amplification	2.4
								MDM2 amplification	2.4
								ETV1 amplification	2.4
								MAP2K4 amplification	2.4
								TP53	45.2
								ATRX	14.3
								RB1	9.5
								PTEN	7.1
								DICER1	4.8
								BRCA2	4.8
								ARID1A	2.4
								PIK3CA	2.4
								RAD51B	
								TSC2	
								ABCG2, ARID2, CDC73, CDKN1C, CYP2D6, FANCA, FBXW7, KMT2C, MRE11A, MYH7, PAH, PALB2, PDE11A, PER1, PMS2, SLCO1B1, UGT1A1	2.4
Chondrosarcoma	12	23-59; 34.5	F=4; M=8	Mesenchymal (8.3%), extracellular myxoid (16.6%)	Sacrum (8.3%), NOS (91.6%)	Lung (8.3%)	TMB-H=1; MSI-H=0; PDL1=0	*HEY1-NCOA2* fusion	8.3
								TP53	16.7
								IDH1	8.3
								NF1	8.3
Osteosarcoma	17	9-47; 17	F=4; M=13	Genic sarcoma (5.9%)	Tibia (29.4%), femur (17.6%), shoulder (5.9%), pelvis (5.9%), nasal cavity (5.9%), humerus (5.9%), fibula (5.9%), NOS (23.5%)	Ribs (5.9%)	TMB-H=1; MSI-H=0; PDL1=1	*PDGFRA* amplification	5.9
								KIT amplification	5.9
								MAP3K1 amplification	5.9
								CDKN1A amplification	5.9
								CCND3 amplification	5.9
								KDR amplification	5.9
								CUL4A amplification	5.9
								TP53	17.6
								RB1	5.9
								AKT1	5.9
								CHEK2	5.9
								ABCG2, CHAT, CHRNG, FAM92A1, GDF6, GJB2, GJB4, HMBS, MVK, NUBPL, PADI3, PRSS56, TBXAS1	5.9
Ewing's sarcoma	11	14-57; 22	F=2; M=9	-	Hip (9.1%), spine (9.1%), soft tissue (9.1%), NOS (72.7%)	-	TMB-H=1; MSI-H=0; PDL1=0	*CDK4* amplification	9.1
								EWSR1-ERG fusion	9.1
								EWSR1-FLI1 Fusion	18.2
								TP53	27.3
								STAG2	9.1
								MTHFR	9.1
								SLCO1B1, ARID1A, EPHB2	9.1

Contd...

SUPPLEMENTARY APPENDIX 7: Contd...

Sarcoma type	Number of patients	Age range; median	Gender	Subtypes	Primary sites	Metastatic sites	TMB-H, MSI-H, PD-L1 +	Genomic alterations	Frequency (%)
Liposarcoma	17	28-73; 54	F=8; M=9	Dedifferentiated (41.2%), pleomorphic (23.5%), myxoid (5.8%)	Retroperitoneal and peritoneal (37.5%), NOS (62.5%)	-	TMB-H=2; MSI-H=0; PDL1=1	*CDK4* amplification	58.9
								MDM2 amplification	53
								DDIT3 amplification	11.7
								FGFR2 amplification	5.8
								FGFR1 amplification	5.8
								YAP1 amplification	5.8
								GLI1 amplification	5.8
								HMGA2 amplification	5.8
								ZC3H7B-BCOR fusion	5.8
								MTHFR	17.4
								TP53	11.7
								FBXW7	5.8
								BRCA2	5.8
								NF1	5.8
								RB1	5.8
								ABCC6	5.8
Angiosarcoma	7	26-74; 57	F=2; M=5	Angiosarcoma squamous cell carcinoma (14.3%), hemangioendothelioma (14.3%)	Scalp (28.6%), buccal mucosa (14.3%), tibia (14.3%), NOS (42.9%)	-	TMB-H=1; MSI-H=0; PDL1=1	*ARID1A*	28.6
								BRAF	14.3
								PIK3CA	14.3
								TP53	14.3
								TERT	14.3
								KEAP1	14.3
								KMT2D, KMT2C, RASA1, SETD2, TSC	14.3
Synovial sarcoma	20	17-70; 33	F=7; M=13	Biphasic (5%)	Femur (5%), lung (10%), thigh (5%)	-	TMB-H=2; MSI-H=0; PDL1=1	*CDK4* amplification	25
								MDM2 amplification	12.5
								PDGFRA amplification	12.5
								ERBB3 amplification	12.5
								CRKL amplification	12.5
								SS18-SSX2 fusion	25
								SS18-SSX1 fusion	12.5
								SS18-SSX3 fusion	12.5
								ASPSCR1-TFE3 fusion	12.5
								C10orf68-CCDC7 fusion	12.5
								MTHFR	12.5
								OLFML2B	12.5
								TSHR	12.5
Rhabdomyosarcoma	9	12-76; 35	F=2; M=7	Pleomorphic (11.1%), alveolar (22.2%)	Prostate (22.2%), uterus (11.1%), lung (11.1%), connective tissue (11.1%), accessory sinuses (11.1%)	-	TMB-H=2; MSI-H=0; PDL1=0	*PAX3-FOXO1* fusion	11.1
								TP53	22.2
								ATM	11.1
								ATRX	11.1
								LATS2	11.1
								MTHFR	11.1
								RB1	11.1
								SLCO1B1	11.1
								TSC2	11.1

Contd...

SUPPLEMENTARY APPENDIX 7: Contd...

Sarcoma type	Number of patients	Age range; median	Gender	Subtypes	Primary sites	Metastatic sites	TMB-H, MSI-H, PD-L1 +	Genomic alterations	Frequency (%)
Spindle cell sarcoma	10	39-73; 55	F=6; M=4	-	Abdominal wall (10%), lung (20%), arm (10%), thigh (10%), uterus (10%), subcutaneous connective tissue (10%)	Lung (20%)	TMB-H=2; MSI-H=0; PDL1=0	*CDK4* amplification	10
								MDM2 amplification	20
								ATRX	10
								ERCC2	10
								MSH3	10
								MTHFR	10
								PIK3CA	10
								TP53	10
								TSC2	10
Soft tissue sarcoma (STS) and sarcoma NOS	41	13-92; 47	F=8; M=13	-	Sarcoma NOS (63.4%), STS NOS (4.9%), abdominal wall (2.4%), testicle (2.4%), lower limb (2.4%), retroperitoneum and peritoneum (12.2%), brain (2.4%), knee (2.4%), parotid gland (2.4%), connective tissue (2.4%), kidney (2.4%)	Lung (2.4%)	TMB-H=10; MSI-H=0; PDL1=0	*SMARCA4* amplification	3
								CDK4 amplification	2
								MDM2 amplification	2
								KEAP1 amplification	2
								BLM amplification	2
								NF2 amplification	1
								RUNX1 amplification	1
								FGFR1 amplification	1
								NOTCH1 amplification	1
								HGMA2 amplification	1
								NCOA3 amplification	1
								APC	12.2
								TP53	12.2
								BCHE	4.9
								BRCA2	4.9
								CHEK2	4.9
								GNAS	4.9
								MTHFR	4.9
								PTEN	4.9
								SLCO1B1	4.9
								STK11	4.9
								EGFR	2.4
								BARD1	2.4
								BCOR	2.4
								ERCC2	2.4
								ABCG2, ALDH7A1, CEP290, CTRC, CYP2D6, DIS3, ECHS1, FANCD2, KRT8, MAP3K1, NUBPL, OCA2, PADI3, PINK1, ROS1, SLC26A4, SLC37A4, SOX2, VKORC1	2.4
GISS	17	40-72; 59	F=7; M=10	-	Rectum (5.9%), NOS (94.1%)	-	TMB-H=1; MSI-H=0; PDL1=0	*BRCA2*	5.9
								KIT	88.2
								PDGFRA	11.8
								PTEN	5.9
								RAD54B	5.9
								RB1	5.9

Contd...

SUPPLEMENTARY APPENDIX 7: Contd...

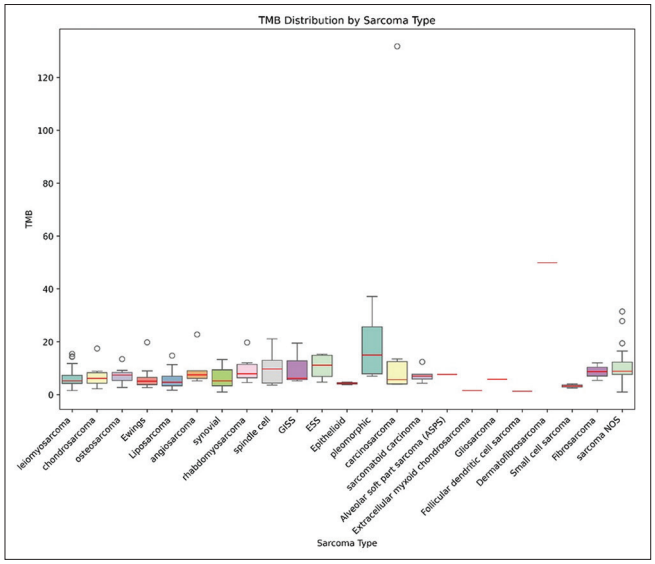
Sarcoma type	Number of patients	Age range; median	Gender	Subtypes	Primary sites	Metastatic sites	TMB-H, MSI-H, PD-L1 +	Genomic alterations	Frequency (%)
ESS	5	44-63; 56	F=5; M=0	-	Ovary (20%), NOS (80%)	Pelvis (20%)	TMB-H=1; MSI-H=0; PDL1=1	*RB1*	40
								TP53	40
								PIK3CA	20
								MSH2	20
								MSH3	20
								PTEN	20
								RAD50	20
								NF1	20
								MEN1, ATRX, ARID1A, FLCN, POLG, SUZ12	20
Epitheloid sarcoma	4	18-55; 29	F=3; M=1	-	Connective tissue (25%), NOS (75%)	-	TMB-H=1; MSI-H=0; PDL1=1	*DNMT3A*	25
								NF2	25
								RET	25
Pleomorphic	4	74-80; 75	F=0; M=4	-	Leg (25%), lung (25%), NOS (50%)	-	TMB-H=1; MSI-H=0; PDL1=0	*KRAS* amplification	25
								SPG7	25
								TP53	50
								RAD50	25
Carcinosarcoma	20	40-75; 57.5	F=20; M=0	-	Ovary (25%), endometrium (45%), thyroid and parathyroid (5%), uterus (10%), lung (5%), breast (5%), gallbladder (5%)	Lung (5%), ovary (5%), gastric shwannoma and liver (5%)	TMB-H=3; MSI-H=0; PDL1=1	*CCND1* amplification	5
								FGF3 amplification	5
								FGF4 amplification	5
								TP53	80
								PIK3CA	20
								PTEN	20
								ARID1A	10
								CTNNB1	10
								CYP2D6	10
								FBXW7	10
								KRAS	10
								PIK3R1	10
								ATM	5
								BCOR	5
								BRCA1	5
								BRCA2	5
								FGFR2	5
								POLD1	5
								RB1	5
								APC, CASP8, CBFB, CDKN2A, KEAP1, KMT2D, MEN1, NOTCH1, NOTCH4, RAD50, SOX17, TSC1, XPO1	5
Sarcomatoid carcinoma	9	21-69; 53	F=2; M=7	-	Liver (11.1%), head-and-neck (11.1%), renal (44.4%), mediastinum (11.1%), lung (11.1%)	Stomach (11.1%)	TMB-H=2; MSI-H=0; PDL1=0	*PDGFRA* amplification	11.1
								TP53	55.5
								NF2	44.4
								ARID1A	22.2
								BCOR	11.1
								KRAS	11.1
								MLH1	11.1
								MSH3	11.1
								NRAS	11.1
								SETD2	11.1
								SMARCA4	11.1

Contd...

SUPPLEMENTARY APPENDIX 7: Contd...

Sarcoma type	Number of patients	Age range; median	Gender	Subtypes	Primary sites	Metastatic sites	TMB-H, MSI-H, PD-L1 +	Genomic alterations	Frequency (%)
								FLCN, JAK1, LIRF, MAGEC3, TSC1	11.1
Alveolar soft part sarcoma (ASPS)	3	21-40; 30	F=1; M=2	-	-	Brain (33.3%)	TMB-H=0; MSI-H=0; PDL1=0	*ASPSCR1-TFE3* fusion	33.3
Gliosarcoma	2	34-53; 43.5	F=0; M=2	-	-	-	TMB-H=0; MSI-H=0; PDL1=1	*TERT*	50
								PTEN	50
								TP53	100
								RB1	50
								TSC2	50
Follicular dendritic cell sarcoma	3	40-53; 52	F=1; M=2	-	-	Lung (33.3%)	TMB-H=0; MSI-H=0; PDL1=0	No clinically significant variants found	-
Dermatofibrosarcoma	2	38-50; 44	F=2; M=0	-	-	-	TMB-H=0; MSI-H=0; PDL1=0	No clinically significant variants found	-
Small-cell sarcoma	2	12-15; 13.5	F=1; M=1	-	Lung (50%), NOS (50%)	-	TMB-H=0; MSI-H=0; PDL1=0	*CTNNB1*	50
								MTHFR	50
Fibrosarcoma	2	50-51; 50.5	F=2; M=0	-	Scalp (50%), breast (50%)	-	TMB-H=1; MSI-H=0; PDL1=0	*CRKL* amplification	50
								MAPK1 amplification	50
								COL1A1-PDGFB fusion	50
								TP53	50
Myofibroblastic sarcoma	2	36-66; 51	F=1; M=1	-	-	-	TMB-H=0; MSI-H=0; PDL1=0	No clinically significant variants found	-
Cyst adenosarcoma	1	64	F=1	-	-	-	TMB-H=0; MSI-H=0; PDL1=0	No clinically significant variants found	-
Fibro myxoid sarcoma	1	42	F=1	-	-	-	TMB-H=1; MSI-H=0; PDL1=0	No clinically significant variants found	-

F=Female, M=Male, TMB-H=High tumor mutation burden, MSI-H=High microsatellite instability, PD-L1 + = PD-L1 expression positive, ESS=Endometrial stromal sarcoma, GISS=Gastrointestinal stromal sarcoma, NOS=Not otherwise specified



SUPPLEMENTARY APPENDIX 8: Box-and-whisker plot of tumor mutational burden (TMB) across the various sarcoma types