OXFORD JOURNAL
OF STUDENT
SCHOLARSHIP

Pan-Cancer Relevance of SMARCE1 as a Potential Cancer Biomarker in Low Grade Glioma and Other Malignancies

Amay Kashyap Deka
amaykashyapdeka@gmail.com

ABSTRACT

Cancer is still one of the top causes of death in the world and is the result of genetic and epigenetic changes that lead to unchecked growth of cells. Existing treatments include surgery, chemotherapy, radiotherapy and targeted therapy, which have enhanced outcomes, but are hampered by drug resistance, tumour heterogeneity, late detection and systemic toxicity. These restrictions underscore the critical need for new biomarkers that can provide improved prognosis and direct targeted treatment.

SMARCE1 (BAF57) is a subunit of the SWI/SNF chromatin-remodelling complex that has been individually associated with invasive progression in breast cancer, resistance to EGFR in lung cancer, and aggressive behaviour in gastric cancer. It has been shown to act as both an oncogene or tumour suppressor, depending on the tissue location. In the present study, we explored the role of SMARCE1 across cancer types using publicly available databases (GEPIA2 and cBioPortal). SMARCE1 expression levels were descriptively assessed in tumour versus normal tissues. In the majority of cancers, SMARCE1 was expressed at higher levels in tumour tissues compared to normal tissues (with the highest changes seen in THYM, DLBC and CHOL and smaller changes in GBM, LGG and PAAD). SMARCE1 was also expressed at lower levels in tumour versus normal in LAML, KICH, ACC and OV. Alterations in SMARCE1 were rare across cancers (less than 1% of patients in the pan-cancer dataset) with the majority of alterations being amplification in epithelial cancers and mutations and deletions in central nervous system cancers. No significant correlation between SMARCE1 alteration and overall survival was observed across any of the cohorts. In low-grade glioma (LGG), patients with alterations in SMARCE1 had a median overall survival of 15.30 months compared to 49.00 months for patients without alterations in SMARCE1 (log-rank $p = 0.0839$). This result is based on a small number of patients and does not adjust for other known prognostic factors. Collectively, alterations and expression of SMARCE1 appear to vary across cancers in a tissue specific manner. LGG is also observed to be a hypothesis generating cancer and further study with adjustment for known prognostic factors and larger patient numbers is needed to assess a potential role for SMARCE1 in prognosis.

Keywords: SMARCE1; BAF57; SWI/SNF chromatin remodelling complex; pan-cancer analysis; tumour biomarker; low-grade glioma; prognosis

Oct 2026
Vol 12. No 4.

INTRODUCTION

Cancer remains one of the leading causes of death worldwide, with an estimated 20 million new cases every year, projected to reach 25 million cases by 2050, and major challenges arising from molecular heterogeneity within as well as across cancer types. Identifying genes whose regulation is preserved across diverse cancers could help us make strides in creating prognostic biomarkers and therapeutic targets with implications across diverse cancer subtypes, thereby reducing dependence on computationally intensive tumour-type-specific characterisation for each candidate gene.

Prognostic biomarkers estimate disease progression, recurrence risk, and treatment urgency, yet several clinically significant cancer subtypes such as kidney renal papillary cell carcinoma (KIRP) and low-grade glioma (LGG) lack reliable molecular prognostic biomarkers. Cancer genomic repositories like The Cancer Genome Atlas (TCGA) and the Genotype-Tissue Expression project (GTEx) enable comprehensive expression, genomic, and clinical analysis at a pan-cancer scale, facilitating biomarker discovery and hypothesis generation prior to clinical validation.

Chromatin remodelling complexes regulate the access of the protein-based transcription machinery to the different parts of the genome, thereby determining which gene programs are active within a cell. Dysfunction of these complexes is a recognised hallmark of cancer, enabling oncogenes to silence tumour suppressor programs while re-activating developmental gene expression. Among them, the SWI/SNF (i.e., BAF) is the most frequently mutated chromatin regulatory complex in cancer (Kadoch et al., 2013) with broad mutation patterns across 20% of all tumours. SWI/SNF activity has also been traced in nucleotide excision and DNA double-strand break repair.

The SWI/SNF complex comprises approximately 29 subunits divided into sub-complexes (cBAF, PBAF, and ncBAF) with unique functions (Kadoch et al., 2013). Previous cancer literature has concentrated on a small set of subunits with known effects, such as the SMARCA4, ARID1A, ARID1B, PBRM1, and ARID2 (Shain & Pollack, 2013). By contrast, many accessory as well as structural subunits remain substantially understudied and examined only in isolated cancer subtypes, if at all, despite being part of the same SWI/SNF complex, whose disruption drives oncogenesis. This exposes a gap in the literature for genes of potential pan-cancer relevance without systematic analysis.

SMARCE1 (SWI/SNF-related, matrix-associated, actin-dependent regulator of chromatin, subfamily E, member 1) is located on chromosome 17q21 and encodes BAF57, a core structural subunit of the SWI/SNF complex that contains a high mobility group (HMG) DNA-binding domain. SMARCE1 upregulation has been associated with gastric carcinoma (Liu et al., 2019) and an invasive progression in breast and ovarian cancers (Sokol et al., 2017). Due to the tissue specific nature of SMARCE1 and absence of a systematic analysis of this gene across cancers, we sought to characterize SMARCE1 expression, mutational and genomic change patterns, mutation distribution and unadjusted survival data across publicly available cancer datasets with the intention of generating testable hypotheses.

MATERIALS AND METHODS

Data Access

Data were accessed through two publicly available platforms, the cBioPortal for Cancer Genomics (https://www.cbioportal.org/) and Gene Expression Profiling Interactive Analysis 2 (GEPIA2) (http://gepia2.cancer-pku.cn/#index). cBioPortal hosts curated genomic and clinical data from published cancer studies (Cerami et al., 2012; Gao et al., 2013; de Bruijn et al., 2023), while GEPIA2 provides uniformly processed RNA sequencing data from The Cancer Genome Atlas (TCGA) and the Genotype-Tissue Expression (GTEx) projects (Tang et al., 2017; Tang et al., 2019). They were used to analyse gene expression, OncoPrint, genetic alteration, somatic mutation distribution and survival of SMARCE1. As all data were de-identified, ethical approval was not required.

Data Selection

Cancer genomics datasets that were publicly available were selected from the cBioPortal for Cancer Genomics platform to examine the contribution of SMARCE1 in various cancer types. The inclusion criteria for the studies to be included in this analysis were that they must contain genomic data (mutations, copy number alterations), clinical and survival data from patients with cancer, be a large-scale study, and be publicly accessible. Studies that called somatic mutations, copy-number changes, and annotation of overall survival were included in the analysis. Studies that were missing any of these three outputs were excluded.

Twelve pan-cancer studies with 148,790 patients and 156,010 cancer samples were included in this analysis. These were the studies catalogued under the pan-cancer and mixed cancer-type categories in cBioPortal that met the criteria above. The subsequent datasets were used for the pan-cancer analysis: Cancer Therapy and Clonal Hematopoiesis (MSK, Nat Genet 2020), China Pan-cancer (OrigiMed, Nature 2022), Metastatic Solid Cancers (UMich, Nature 2017), MSK MetTropism (MSK, Cell 2021), MSK-CHORD (MSK, Nature 2024), MSK-IMPACT 50K Clinical Sequencing Cohort (MSK, Cancer Cell 2026; Bandlamudi et al., 2026), MSK-IMPACT Clinical Sequencing Cohort (MSK, Nat Med 2017), MSS Mixed Solid Tumors (Broad/Dana-Farber, Nat Genet 2018), Pan-cancer Analysis of Whole Genomes (ICGC/TCGA, Nature 2020), SUMMIT-Neratinib Basket Study (Multi-Institute, Nature 2018), TMB and Immunotherapy (MSK, Nat Genet 2019) and Tumors with TRK Fusions (MSK, Clin Cancer Res 2020).

As the gene SMARCE1 has been described previously in breast, lung, gastric and central nervous system tumours, five tumour-focused query sets were also constructed using the same criteria, by selecting every study listed under the corresponding cancer-type category in cBioPortal (breast, lung, CNS/brain, oesophagogastric and glioma) that met the inclusion criteria. These included 36 studies on breast cancer, 41 studies on lung cancer, 33 studies on central nervous system (CNS) and brain tumours, 19 studies on oesophagogastric cancer and 5 studies on low-grade gliomas. Some of these studies used more than one sample per patient. Hence, the number of patients and the number of cancer samples were reported separately. Each study is treated independently by cBioPortal and records are not merged between studies. Therefore patients can occur in more than one study in a query set (for example TCGA patients can be in the original TCGA study and also in the TCGA PanCancer Atlas release) and can further appear in more than one query set. For instance there are 5 low grade glioma studies which are a subset of the 33

CNS/brain tumour studies. Furthermore some TCGA and MSK studies also occur in the pan-cancer studies. For this reason the 6 query sets should be considered overlapping views of the available data rather than independent replication sets. The numbers of patients and samples in Table 1 are the numbers of records as returned by cBioPortal for each query. All queries can be performed using the "Query by Gene" module in cBioPortal by using the gene symbol SMARCE1 and the list of studies detailed above.

The composition of each of the six query sets is summarised in Table 1.

Table 1. Composition of the cBioPortal query sets analysed in this study.

Query set	Studies	Samples	Patients	Patients altered	Samples altered
Pan-cancer	12	156,010	148,790	356 (<1%)	362 (<1%)
Breast cancer	36	17,758	16,005	485 (3%)	501 (3%)
Lung cancer	41	15,910	13,241	112 (<1%)	112 (<1%)
CNS / brain tumours	33	10,009	9,038	38 (<1%)	40 (<1%)
Oesophagogastric cancer	19	4,909	4,810	268 (6%)	270 (6%)
Low-grade glioma	5	2,258	2,204	4 (<1%)	4 (<1%)

Query sets overlap: the low-grade glioma studies are contained within the CNS/brain set, and TCGA and MSK studies are shared between the pan-cancer and tumour-focused sets.

Gene Expression Analysis

The GEPIA2 web server, which utilises RNA sequencing data from the TCGA and GTEx projects to normalise gene expression levels, was used to analyse the expression of SMARCE1. The gene symbol SMARCE1 was submitted to the GEPIA2 server, and its expression levels were compared between tumour and normal tissues for each of the 31 cancer types included in the TCGA database. The normalisation method used by GEPIA2 was employed as the default method for normalising expression levels: expression was calculated as transcripts per million (TPM) and log2(TPM + 1) transformed. The median TPM levels of SMARCE1 expression for each cancer type were plotted as paired bars. Mesothelioma (MESO) and uveal melanoma (UVM) were not included in these analyses due to the lack of normal tissue samples for these cancers in the TCGA database. The bar plot in GEPIA2 only displays median TPM values and there is no statistical analysis between tumour and normal samples. No further statistical analysis or multiple comparison correction was performed.

OncoPrint Analysis

The cBioPortal for Cancer Genomics was used to analyse and visualise SMARCE1 genomic alteration data across all cancers. The gene symbol "SMARCE1" was entered under the "Query by Gene" section to fetch alteration information from all selected studies, and the profile was plotted in an OncoPrint in which each column was one patient sample and each row was one gene. Missense, truncating, inframe and splice mutations, structural variants, amplifications and deep deletions were represented by colour codes

generated automatically by the platform. A sample with at least one alteration was classified as "altered" and all others as "unaltered".SMARCE1 mutation types were combined into an "altered" category since SMARCE1 altered is rare and the number of samples with any one type of alteration was too small to perform analysis across all cohorts.

Analysis of Distribution of Somatic Mutations

The distribution of somatic mutations along the SMARCE1 protein was examined using the cBioPortal platform. The built-in module for visualizing mutation data was used to produce a lollipop plot of the mutations along the length of the SMARCE1 protein (411 amino acids in length, encoded by transcript NM_003079). The mutation types (missense, truncating, splice-site and inframe mutations) and whether the mutations were classified as putative drivers or variants of unknown significance were annotated on the plot. Further annotations regarding the cancer hotspots, post-translational modifications of the protein (dbPTM), and the annotation of the exons containing the protein sequence also appear on the plot. Additionally, the position of the HMG box domain within the protein is also represented on the plot.

Analysis of Survival Prognosis

The "Comparison/Survival" module of cBioPortal was used to perform survival analysis. Patients were divided into an "altered" and an "unaltered" group according to the genomic alterations detected in their tumour samples, patients whose tumour samples had not been genomically profiled were excluded from the analysis, and there was a requirement of at least three patients in each group. Depending on the annotation available for each cancer type, five different survival endpoints were calculated for each group: overall, disease-free, disease-specific, progression-free and central nervous system progression-free survival. The log-rank test was used in the platform to assess the statistical significance of the calculated survival endpoints, with p-values < 0.05 considered statistically significant, and false discovery rate corrected q-values reported where more than one survival endpoint was calculated. Unadjusted two-group comparisons were used (as in cBioPortal) as multivariable modelling is not available within the platform. Covariates including age, tumour grade, IDH mutation status and 1p/19q co-deletion status were not included. Thus, the survival analyses presented represent simple associations and may be confounded.

Genetic Alteration Analysis

The "Cancer Types Summary" module on cBioPortal was used to analyse the genetic alterations within the cancer types studied. Each of the cancer types was represented in the pan-cancer genetic alteration study, as well as each of the tumour types represented within the five tumour-focused study data sets. Each of these cancer types was filtered to contain at least 10 patients with sequenced tumour genomes, and the types of genetic alterations within each cancer type were represented as individual bars within the study data sets.

RESULTS

Gene Expression Analysis

The analysis of the expression of SMARCE1 showed that SMARCE1 is expressed at a higher median level in tumour than normal tissue in most of the cancer types examined (Figure 1). The increase in

expression is most prominent in cancers of the thymoma (THYM), diffuse large B-cell lymphoma (DLBC) and cholangiocarcinoma (CHOL) cancers, with a similar (but to a lesser extent) increase seen in glioblastoma multiforme (GBM), brain lower grade glioma (LGG) and pancreatic adenocarcinoma (PAAD). However, SMARCE1 is expressed at lower levels in tumours of acute myeloid leukaemia (LAML), kidney chromophobe (KICH), adrenocortical carcinoma (ACC) and ovarian serous cystadenocarcinoma (OV). These are descriptive differences in median TPM. Thus, the pattern of SMARCE1 expression appears to be dependent upon the tissue of origin.

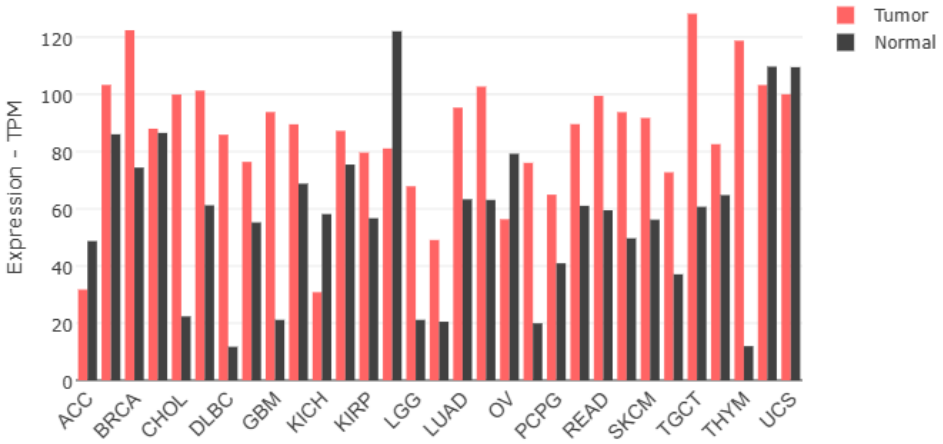

Fig. 1. Gene expression analysis. The bar graph indicates median SMARCE1 expression (TPM) in tumour tissues and their corresponding normal tissue for the 31 TCGA cancer types with matched TCGA/GTEx normal samples. Generated with the GEPIA2 expression bar-plot module and reproduced without modification.

OncoPrint Analysis

Genomic alterations in SMARCE1 were assessed across the 12 pan-cancer datasets using OncoPrint visualisation (Figure 2). Alterations were rare, occurring in 356 of 148,790 patients and 362 of 156,010 samples, which is less than 1% in both cases. The great majority of variants were amplifications, with deep deletions constituting a smaller second group; missense, truncating, inframe and splice mutations of unknown significance were less common.

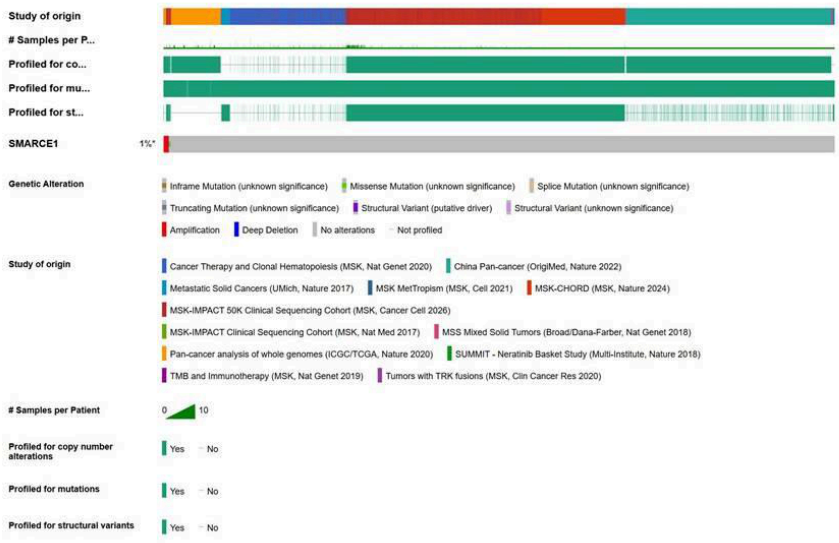

Fig. 2. OncoPrint analysis. Genomic alterations in SMARCE1 across the 12 pan-cancer datasets (pan-cancer query set, Table 1). OncoPrint generated in cBioPortal; screenshot reproduced without modification.

Genetic Alteration Analysis

The genetic alteration status of SMARCE1 in different tumour types within the pan-cancer query set was analysed. Figure 3 indicates that the highest alteration frequency appears in embryonal tumours, at about 12% of patients, followed by oesophagogastric cancer at about 6.6%, medulloblastoma at about 4.7% and uterine endometrioid carcinoma at about 4.1%, with several gastrointestinal and epithelial cancers between roughly 2% and 3%. Amplification was the primary alteration type in almost every category and mutation the secondary type. Uterine endometrioid carcinoma was the clear exception, its bar being made up almost entirely of deep deletions, which suggests that SMARCE1 is gained at the copy-number level in some epithelial tumours and lost in others.

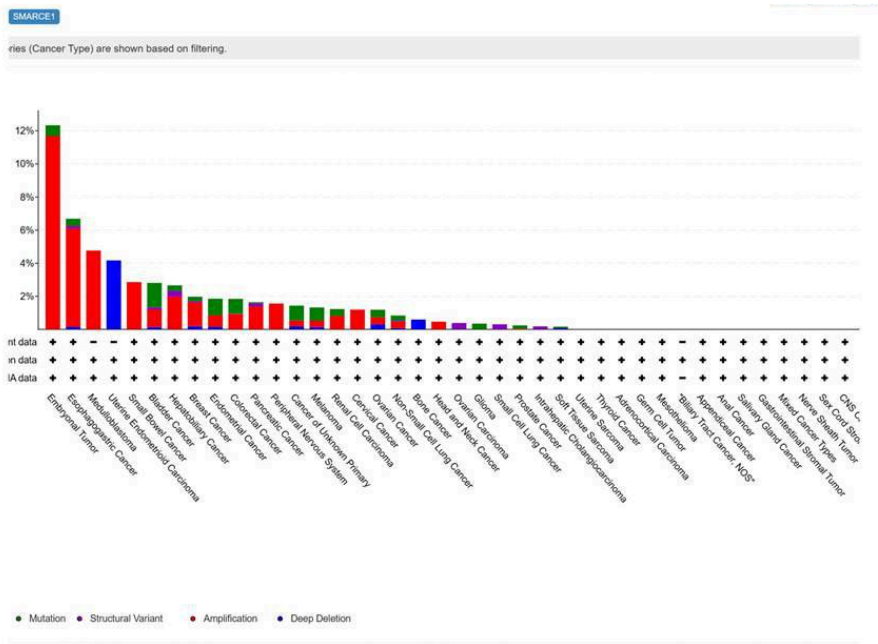

Fig. 3. Genetic alteration analysis. The bar graph indicates the alteration frequency of SMARCE1 in different cancer types within the pan-cancer query set (Table 1). Generated with the cBioPortal "Cancer Types Summary" module; screenshot reproduced without modification.

Analysis of Distribution of Somatic Mutations in SMARCE1

The lollipop plot of SMARCE1 shows that the somatic mutations are spread throughout the gene and not clustered in any one part of the gene (Figure 4). The Cancer Hotspots track is empty for this gene. Most of the amino acids are changed in only one patient with the most common change being the change at the splice-site at codon 53 (X53_splice) in four patients. There is a modest increase in the density of mutations in the region that contains the HMG box (residues 70 to 133). Some of the mutations fall near amino acids that can be phosphorylated or otherwise modified by post-translational modifications. The pattern of mutations is typical of a gene that is inactivated in multiple different ways rather than one activated by a recurrent change.

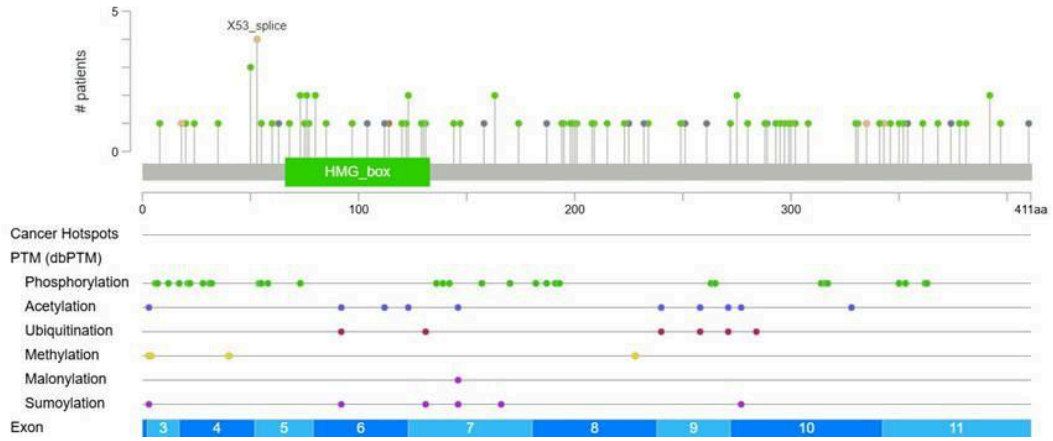

Fig. 4. Distribution of somatic mutations in SMARCE1. The lollipop plot illustrates the distribution of somatic mutations along the protein sequence in the pan-cancer query set. Generated with the cBioPortal "Mutations" module; screenshot reproduced without modifications.

Survival Prognosis in the Pan-Cancer Cohort

Figure 5 shows the probability of overall survival in the altered and unaltered groups, for which data were available for 14,011 patients. There were 212 altered patients with 40 deaths and 13,799 unaltered patients with 2,688 deaths, and the median overall survival was 48.62 months (95% CI: 40.50–NE) and 59.34 months (95% CI: 52.17–80.71) respectively, with a log-rank p-value of 0.975. Pooling amplifications with deletions across tumour types with very different baseline survival would be expected to cancel out any effect specific to one setting, so the analysis was repeated in each of the five tumour-focused query sets.

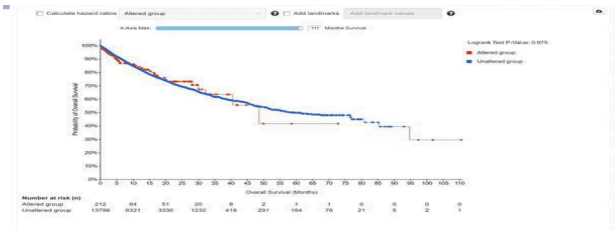

Fig. 5. Analysis of survival prognosis. The Kaplan-Meier plot indicates probability of overall survival in the pan-cancer cohort (unadjusted log-rank comparison). Generated with the cBioPortal "Comparison/Survival" module; screenshot reproduced without modification.

Role of SMARCE1 in Brain Tumours and Low-Grade Glioma

SMARCE1 was altered in 38 of 9,038 patients in the CNS cohort, the lowest of all six query sets. What stands out is the type of alteration, because only missense, splice and truncating mutations together with deep deletions were recorded and not a single amplification was found. Figure 6 indicates that the highest frequency appears in diffuse glioma, at about 5.4% and 4.7% in the two GLASS Consortium cohorts, followed by medulloblastoma at about 2.2%, all composed almost entirely of mutations. Figure 7 shows

that 10 altered patients had a median overall survival of 15.31 months against 24.77 months (95% CI: 23.93–25.69) in 6,462 unaltered patients, which was not significant (log-rank p = 0.119).

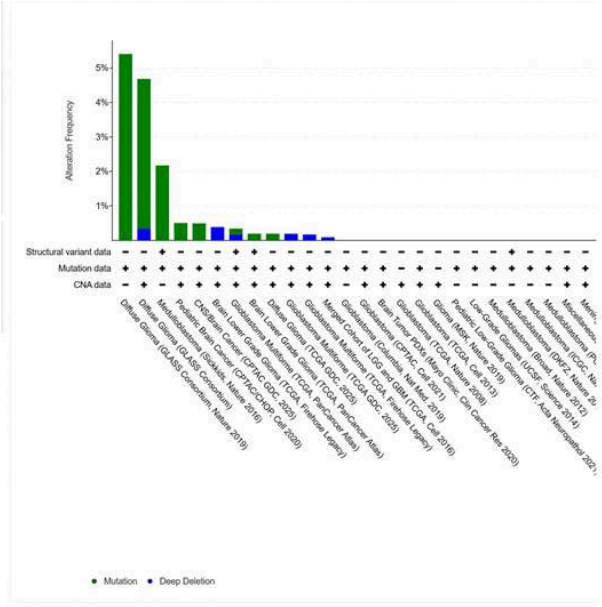

Fig. 6. Genetic alteration analysis. The bar graph indicates the alteration frequency of SMARCE1 in brain and CNS related tumour types (CNS/brain query set, Table 1). Generated with the cBioPortal "Cancer Types Summary" module; screenshot reproduced without modification.

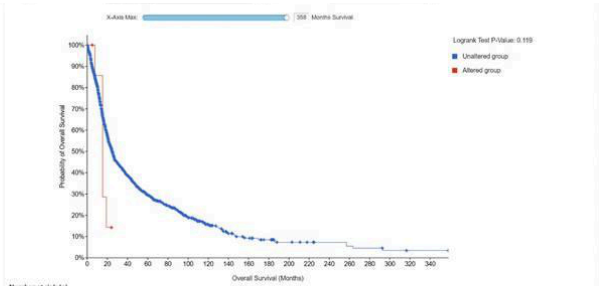

Fig. 7. Analysis of survival prognosis. The Kaplan-Meier plot indicates probability of overall survival in CNS and brain tumours (CNS/brain query set; unadjusted log-rank comparison). Generated with the cBioPortal "Comparison/Survival" module; screenshot reproduced without modification.

A separate set of five low-grade glioma studies, all of which are also contained within the CNS/brain query set, was then analysed. SMARCE1 was altered in only 4 of 2,204 patients, through deep deletions and a single truncating mutation, W104*, in sample TCGA-HT-7601-01 (Figure 8). Residue 104 lies inside the HMG box, so a stop codon here would be predicted to remove the DNA-binding domain and the whole C-terminal half of the protein. Figure 9 shows that these 4 patients had a median overall survival of 15.30 months (95% CI: 15.30–NE) against 49.00 months (95% CI: 44.00–54.77) in 2,078 unaltered patients, the largest separation seen in any query set, although the log-rank test gave a p-value

of 0.0839. With only four altered patients The comparison lacks power and is not adjusted for age, tumour grade, IDH mutation status or 1p/19q co-deletion, all of which are prognostically important in low grade glioma. Thus, while the low grade glioma result is the most interesting aspect of this study, it is also preliminary and hypothesis generating and not indicative of a prognostic impact.

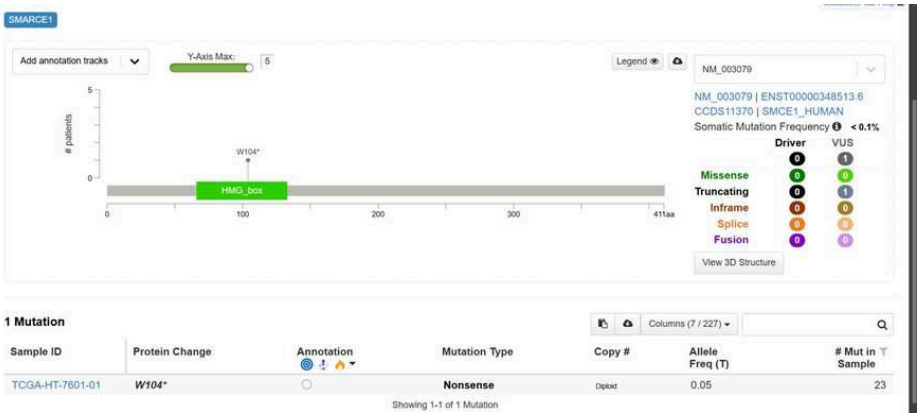

Fig. 8. Somatic mutation of SMARCE1 in low-grade glioma. The single W104* nonsense variant identified in this cohort (low-grade glioma query set, Table 1). Generated with the cBioPortal "Mutations" module; screenshot reproduced without modification.

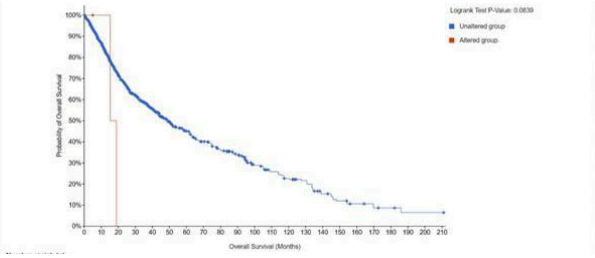

Fig. 9. Analysis of survival prognosis. The Kaplan-Meier plot indicates probability of overall survival in low-grade glioma (four altered versus 2,078 unaltered patients; unadjusted log-rank comparison). Generated with the cBioPortal "Comparison/Survival" module; screenshot reproduced without modification.

Role of SMARCE1 in Breast Cancer

Across 36 breast cancer studies, SMARCE1 was altered in 485 of 16,005 patients and 501 of 17,758 samples, about 3%, and amplification was by far the most common event. Figure 10 indicates that the highest alteration frequency appears in the metastatic and advanced cohorts, reaching about 43% in The Metastatic Breast Cancer Project (Provisional, December 2021) and about 17% in Adenoid Cystic Carcinoma of the Breast (MSK, J Pathol 2015), while the three TCGA Breast Invasive Carcinoma cohorts fell between about 4% and 5.5%.

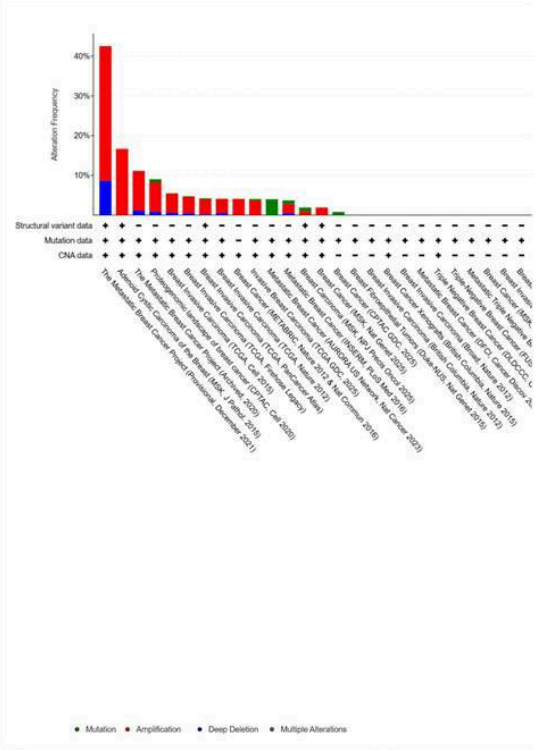

Fig. 10. Genetic alteration analysis. The bar graph indicates the alteration frequency of SMARCE1 in different breast cancer types (breast cancer query set, Table 1). Generated with the cBioPortal "Cancer Types Summary" module; screenshot reproduced without modification.

Figure 11 shows the probability of overall survival. The altered group contained 296 patients with 77 deaths and the unaltered group 6,684 with 1,808 deaths, and the median overall survival was 168.20 months (95% CI: 146.39–196.87) and 148.57 months (95% CI: 140.77–157.80), which was not significant (log-rank p = 0.496). Of the five endpoints tested, however, disease-free survival differed clearly between the groups in 3,653 patients ($p = 3.146 \times 10^{-4}$, $q = 1.573 \times 10^{-3}$), the only endpoint in any query set to remain significant after correction for multiple testing. The disease-free curve itself was not examined here, so the direction of the effect cannot be stated, but SMARCE1 alteration in breast cancer appears to be associated with recurrence rather than with overall survival, which fits work linking the gene to invasion and early metastatic spread (Sokol et al., 2017).

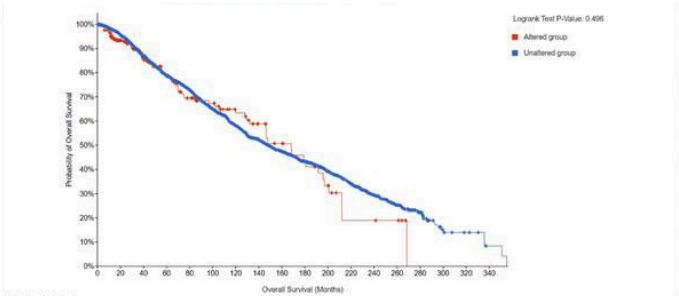

Fig. 11. Analysis of survival prognosis. The Kaplan-Meier plot indicates probability of overall survival in breast cancer (breast cancer query set; unadjusted log-rank comparison). Generated with the cBioPortal "Comparison/Survival" module; screenshot reproduced without modification.

Role of SMARCE1 in Lung and Oesophagogastric Cancer

Across 41 lung cancer studies, SMARCE1 was altered in 112 of 13,241 patients, less than 1%. Figure 12 indicates that the highest frequency appears in Non-Small Cell Lung Cancer (MSK, 2026) at about 11.4% and Non-Small Cell Lung Cancer Brain Metastasis (MSK, Nat Commun 2023) at about 8.4%, with deep deletion making up about 2.2% and 4.2% of these two cohorts and the third-ranked study, Metastatic Non-Small Cell Lung Cancer (MSK, Nature Medicine 2022), consisting of deep deletions alone. Deletion therefore contributes more here than in the other epithelial tumours examined, which is consistent with earlier work showing that loss of SMARCE1 raises EGFR expression and confers resistance to MET and ALK inhibitors (Papadakis et al., 2015). Figure 13 shows that the 81 altered patients had a median overall survival of 49.01 months against 53.29 months in 5,014 unaltered patients (log-rank p = 0.115).

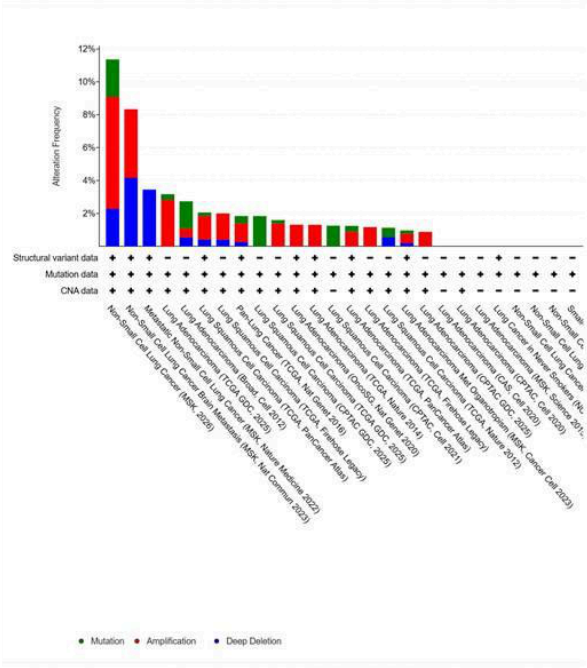

Fig. 12. Genetic alteration analysis. The bar graph indicates the alteration frequency of SMARCE1 in different lung cancer types (lung cancer query set, Table 1). Generated with the cBioPortal "Cancer Types Summary" module; screenshot reproduced without modification.

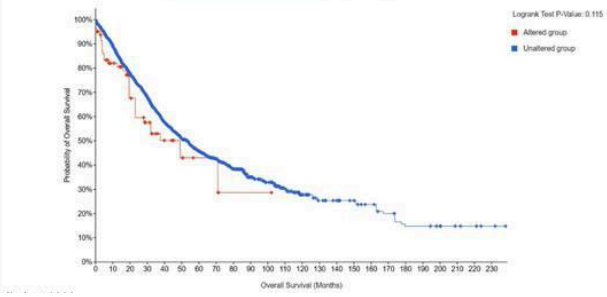

Fig. 13. Analysis of survival prognosis. The Kaplan-Meier plot indicates probability of overall survival in lung cancer (lung cancer query set; unadjusted log-rank comparison). Generated with the cBioPortal "Comparison/Survival" module; screenshot reproduced without modification.

The oesophagogastric set gave the highest alteration frequency of all six, with SMARCE1 altered in 268 of 4,810 patients, about 6%, almost entirely through amplification. Figure 14 indicates that the highest frequency appears in the two Memorial Sloan Kettering cohorts, at about 20.5% and 20.0%, with Esophageal Adenocarcinoma (TCGA, PanCancer Atlas) reaching about 10.4%, and the consistency of this signal across nineteen studies suggests it is a genuine feature of the tumour type. Figure 15 shows that the 250 altered patients had a median overall survival of 46.16 months against 30.88 months in 2,568 unaltered patients, the opposite of what would be expected if amplification were driving aggressive

disease, but the log-rank p-value of 0.709 means no conclusion can be drawn. Table 2 summarises the survival results from all six query sets.

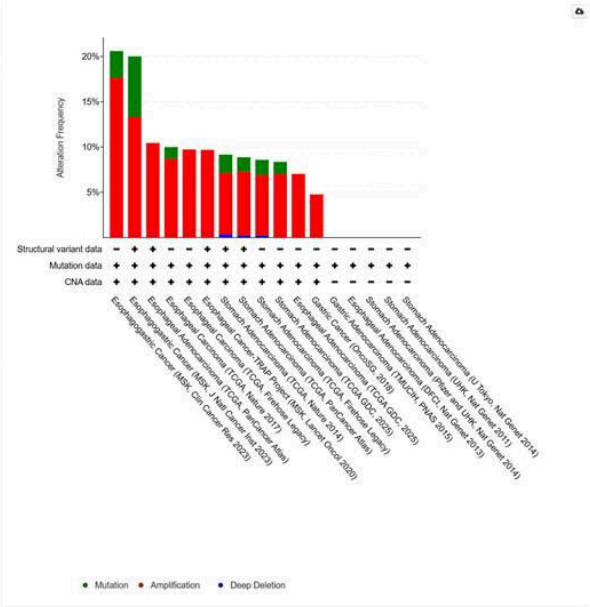

Fig. 14. Genetic alteration analysis. The bar graph indicates the alteration frequency of SMARCE1 in oesophagogastric cancer types (oesophagogastric query set, Table 1). Generated with the cBioPortal "Cancer Types Summary" module; screenshot reproduced without modification.

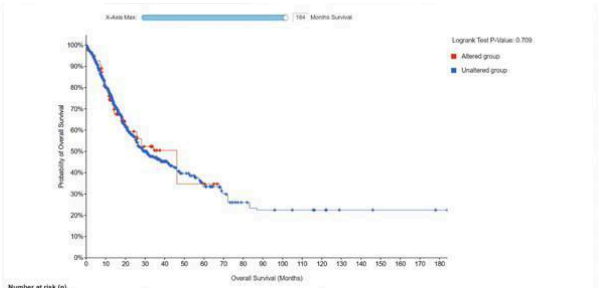

Fig. 15. Analysis of survival prognosis. The Kaplan-Meier plot indicates probability of overall survival in oesophagogastric cancer (oesophagogastric query set; unadjusted log-rank comparison). Generated with the cBioPortal "Comparison/Survival" module; screenshot reproduced without modification.

Table 2. Summary of overall survival outcomes across the six query sets (unadjusted log-rank comparisons).

Query set	Altered (n / events)	Median OS, altered, months (95% CI)	Unaltered (n / events)	Median OS, unaltered, months (95% CI)	Log-rank p
Pan-cancer	212 / 40	48.62 (40.50–NE)	13,799 / 2,688	59.34 (52.17–80.71)	0.975
Breast cancer	296 / 77	168.20 (146.39–196.87)	6,684 / 1,808	148.57 (140.77–157.80)	0.496
Lung cancer	81 / 34	49.01 (23.10–NE)	5,014 / 1,864	53.29 (49.01–55.68)	0.115
CNS / brain tumours†	10 / 6	15.31 (15.30–NE)	6,462 / 3,587	24.77 (23.93–25.69)	0.119
Oesophagogastric cancer	250 / 94	46.16 (25.78–NE)	2,568 / 958	30.88 (26.45–34.29)	0.709
Low-grade glioma†	4 / 2	15.30 (15.30–NE)	2,078 / 776	49.00 (44.00–54.77)	0.0839

OS, overall survival; CI, confidence interval; NE, upper confidence limit not estimable because too few events occurred in the altered group. † Altered group of fewer than 15 patients: these comparisons are severely underpowered and should be interpreted with particular caution. All comparisons are unadjusted for clinical covariates; n and events refer to patients with overall survival data, which is fewer than the total number of patients in each query set (Table 1).

DISCUSSION

SMARCE1, also known as BAF57, encodes the structural subunit of the SWI/SNF complex that contains the high mobility group box that allows it to bind to DNA. The SWI/SNF complex is the most frequently mutated chromatin regulatory complex in human cancer, with alterations in around a fifth of all tumors (Kadoch et al., 2013; Shain & Pollack, 2013). Thus, the alteration of any one subunit of the complex can have different effects in different types of tissue. There are two types of data observed in this analysis that must be distinguished. Expression level data is derived from RNA sequencing data available in GEPIA2 and genomic alteration (mutation or copy number change) is derived from DNA sequencing data available in cBioPortal. Alteration in one does not necessarily imply alteration in the other. For example, the elevated expression of SMARCE1 seen in the majority of tumour types is consistent with reports from gastric cancer, where high levels of SMARCE1 have been linked with deep invasion, lymph node involvement, and poor survival rates (Liu et al., 2019), whereas the loss-of-function mutations and deletions observed here in nervous system tumours are consistent with the tumour-suppressor role

described for SMARCE1 in meningioma (Smith et al., 2013). Thus, SMARCE1 may plausibly act as either an oncogene or a tumor suppressor gene, depending upon the type of tumor that is being examined.

The most informative result of the analysis of the type of alterations in the region was actually not the frequency of the alterations altogether, but rather in how the type of alteration appears to differ between tumour types. For breast and oesophagogastric cancers, the alterations were almost entirely amplifications. These two cancer types also had the highest frequencies of such alterations. Furthermore, SMARCE1 is located on chromosome 17q, the same region as ERBB2 (HER2). Co-amplification of SMARCE1 with HER2 has been described in gastric cancer (Pries et al., 2023). As both of these cancer types contain frequent amplifications of HER2, one hypothesis is that the amplification of SMARCE1 is a passenger event arising because the gene lies within the ERBB2 amplicon of chromosome 17q. ERBB2 copy number was not assessed on the same tumour samples and thus whether SMARCE1 is selected independently of ERBB2 could not be determined. SMARCE1 amplification in these tumour types should not currently be considered a driver event until independent validation of this is performed.

The two cohorts of the central nervous system behaved in the opposite way to the other organs: there were no amplifications at all in the nervous system, all alterations being either mutations or deep deletions. This is the expected behaviour for a tumour suppressor gene, and it agrees with existing work in the field. For example, loss-of-function mutations in the germline lead to spinal and cranial clear cell meningiomas, cancers that are characterized by a complete loss of SMARCE1 protein within the cancer cells (Smith et al., 2013; Smith et al., 2014). Furthermore, loss of SMARCE1 leads to the shift of complexes from the canonical BAF complex to BRD9-containing non-canonical complexes, a change that can be targeted with drugs in cancer cells (St Pierre et al., 2022). A stop in a HMG box in the W104* mutant in a patient with a low grade glioma would result in loss of the DNA binding domain and is thus predicted to be of the same type of loss. The meningioma data however represents a different tumour and a germline mutation. A single truncating mutation in one disease does not demonstrate a role for SMARCE1 loss in low grade glioma or the survival difference seen in the cohort of meningioma patients. This finding should be considered as a hypothesis for further investigation. Consequently, while the comparison of survival between altered and unaltered patients was not statistically significant in either the low-grade glioma or the CNS cohort (as there were only four and ten altered patients, respectively), it is still of some interest that both cohorts exhibited the same tendency for survival times, while recognising that the two cohorts overlap and that neither comparison was adjusted for clinical covariates.

Interpretation of the survivals in this analysis needs to be done carefully for two reasons. Firstly, the analysis is unadjusted. There are a number of well-documented factors that strongly predict survival in low grade gliomas including tumour grade, IDH mutation status, 1p/19q co-deletion and age. IDH wild type tumour in particular have significantly poorer outcomes. If any of these factors were over-represented in the four SMARCE1 altered samples then the difference in median survival could be purely due to confounding. cBioPortal does not offer multivariable survival analysis. Indeed with only four altered samples it would not have been appropriate to perform such an analysis anyway. This option was not explored in this analysis. Secondly, alterations are grouped together as altered. This includes both deletions, amplifications and sequence alterations which are likely to impact SMARCE1 in different ways. This was done as alterations are rare and alteration-specific analyses would not have had sufficient

samples to perform. This is least likely to cause issues in the low grade glioma and CNS analyses as there were no amplifications and all alterations were deletions or mutations. It is more likely to have caused issues in the epithelial data where class specific effects may have been drowned out by this grouping.

In lung cancer the pattern was intermediate, with deletions making up a larger share than in the other epithelial cohorts, and since loss of SMARCE1 confers resistance to MET and ALK inhibitors (Papadakis et al., 2015), deletion may be more useful as a predictive marker of treatment response than as a prognostic marker of survival. Although SMARCE1 is altered in only a small percentage of tumours overall, and its mutations are scattered along the protein without any hotspot, the class of alteration was far from uniform: it is altered mainly through amplification in epithelial tumours and through mutation and deletion in the nervous system. For genes of this type, the class of alteration and the tissue in which it occurs appear to carry more information than the overall frequency does.

LIMITATIONS

Although this study provides a pan-cancer evaluation of SMARCE1, there are some limitations to this study that should be considered. The analyses were performed based upon data from TCGA and other public databases, but there is a high level of heterogeneity between these datasets. For instance, the patients included in these databases were sequenced in different laboratories, with different bioinformatic processing pipelines, and came from different demographic populations altogether.

The size of some of the groups with altered SMARCE1 sequences in relation to survival analyses is particularly limited. For instance, in the cases of low-grade glioma and the CNS cancer cohort, there were only four and ten patients with altered SMARCE1, respectively. Thus, there is little power in these analyses to draw meaningful conclusions from such small sample sizes. Potential confounding by grade, IDH status, 1p/19q co-deletion, age, other prognostic factors or by SMARCE1 itself cannot be excluded due to the lack of adjustment in the survival analyses.Furthermore, alterations to SMARCE1 are grouped together (amplifications, deep deletions and sequence mutations), which may have opposite effects upon SMARCE1 function, and alteration-specific analyses were not possible because of small numbers. The six question sets overlap on patients and studies and are not independent validation sets. No statistical analysis was performed on the gene expression differences. Differences were purely descriptive.Finally, because SMARCE1 is located near the gene ERBB2 on chromosome 17q, some of the amplifications of SMARCE1 in breast and oesophagogastric cancers may be “passenger” events resulting from amplification of the ERBB2 gene instead; this remains a hypothesis that the present analysis could not test.

CONCLUSION

To understand the biological and clinical significance of SMARCE1, its expression and genomic alteration profiles were analysed across a range of cancer types. SMARCE1 expression was higher in tumour tissue in the majority of cancer types, with the largest increases in thymoma, diffuse large B-cell lymphoma and cholangiocarcinoma, but was lower in cancers such as acute myeloid leukaemia, kidney chromophobe, adrenocortical carcinoma and ovarian cancer. Thus, the pattern of SMARCE1 expression

appears to depend upon the tissue in which it is acting, although these differences were descriptive and were not tested statistically.

SMARCE1 was found to be altered in less than 1% of patients with any type of cancer. However, the alterations to SMARCE1 were distributed unevenly across cancer types; there were alterations present in around 12% of patients with embryonal tumours, around 6% of patients with oesophagogastric cancer, and as high as 43% of patients in a metastatic breast cancer cohort. Furthermore, most alterations to SMARCE1 were amplifications of the gene in breast and oesophageal cancer, deep deletions of the gene in lung cancer, but no amplifications of the gene in CNS cancers. Additionally, the mutations to SMARCE1 that scattered throughout the gene were not targeted to any specific base pair within the gene; there were no hotspots for mutations to SMARCE1, and the splice and truncating mutations that were present were of the type that could lead to a loss of gene function.

This analysis does not provide data regarding the independent prognostic value of SMARCE1 beyond known molecular and clinical variables due to the unadjusted nature of the comparisons. However, there was a difference between the survival of those with low-grade glioma according to SMARCE1 status: 15.30 months for those with the altered gene compared to 49.00 months for those without the altered gene, based on only four altered patients and not reaching statistical significance (log-rank $p = 0.0839$). For the CNS cancers as a whole the difference pointed the same way but was smaller, at 15.31 months for those with the altered SMARCE1 compared to 24.77 months for those without the altered gene, though the groups with the altered gene were very small. SMARCE1 appears, therefore, to be a gene that is gained at the copy-number level in some epithelial tumours and that is lost in tumours of the nervous system. The low-grade glioma observation, together with the disease-free survival signal in breast cancer, identifies settings in which a prognostic role could be tested. These observations are preliminary, hypothesis-generating, and require validation in larger cohorts.

FUTURE PERSPECTIVES

The first priority would be to test this signal in a larger cohort of patients with low-grade gliomas. While the difference in survival times between the two groups was quite large, it was based upon only four patients with altered SMARCE1. To test for statistical significance of any observed effect, it would be necessary to perform the test with a sufficient number of patients with altered SMARCE1. Such a study could employ different glioma patient cohorts, or it could employ sequencing of a panel of genes in a large number of glioma patients. Furthermore, the study would have to account for factors like age, tumor grade, IDH gene status, and the presence of co-deletion of the 1p and 19q chromosomes, all of which are known factors that influence the survival time of patients with low-grade gliomas.

A second priority is to separate the effects of amplification and deletion rather than analysing them together. As their effects are likely to be opposite in direction, SMARCE1-amplified and SMARCE1-deleted tumours should be considered separately in future analyses. Within breast and oesophagogastric cancer, SMARCE1 amplifications could potentially occur independently of amplifications in the closely-located ERBB2 gene; examining the relationship between SMARCE1 copy

number and HER2 status in the same tumours will reveal whether SMARCE1 amplification occurs independently of HER2 amplification.

From a clinical point of view, the most promising direction may not be prognosis but prediction of treatment response. Loss of SMARCE1 confers resistance to MET and ALK inhibitors in non-small cell lung cancer, and its loss in meningioma creates a dependency on non-canonical BAF complexes that can be targeted. Studies testing whether SMARCE1 status predicts response to targeted agents would make better use of what is already known about the gene. Integration of proteomic and transcriptomic data would additionally help identify interacting partners and tumour-specific regulatory networks associated with SMARCE1, and combining SMARCE1 status with other molecular and clinicopathological markers may improve patient stratification more than SMARCE1 alone.

ACKNOWLEDGEMENTS

1) I am grateful to my supervising teacher, Ms. Nandini Dutta, Biology Teacher at Delhi Public School, Guwahati, for her constant support, guidance and encouragement throughout my research journey.
2) I am grateful to the principal and faculty of Delhi Public School, Guwahati for checking in with me throughout my research journey and for being supportive of my endeavours.
3) I am grateful to the teams behind cBioPortal for Cancer Genomics, GEPIA2, The Cancer Genome Atlas and the Genotype-Tissue Expression project for making the data and analytical resources which made this project possible openly available.
4) I am grateful to my friends and family for motivating me and supporting me throughout my research journey.

Author's contributions: Amay Kashyap Deka conceptualised the study, carried out the literature survey and the bioinformatics analysis, analysed the data, and wrote and corrected the manuscript.

Conflict of interest: There is no potential conflict of interest to disclose.

Financial support: There is no funding information to disclose.

REFERENCES

Bandlamudi, C., et al. (2026). Cancer type-specific variation in patterns of driver alterations across 50,000 tumors. Cancer Cell, 44(5), 1045–1062.e6. https://doi.org/10.1016/j.ccell.2026.03.003

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., Jacobsen, A., Byrne, C. J., Heuer, M. L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A. P., Sander, C., & Schultz, N. (2012). The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. Cancer Discovery, 2(5), 401–404. https://doi.org/10.1158/2159-8290.CD-12-0095

de Bruijn, I., Kundra, R., Mastrogiacomo, B., Tran, T. N., Sikina, L., Mazor, T., Li, X., Ochoa, A., Zhao, G., Lai, B., ... Schultz, N. (2023). Analysis and visualization of longitudinal genomic and clinical data

from the AACR Project GENIE Biopharma Collaborative in cBioPortal. Cancer Research, 83(23), 3861–3867. https://doi.org/10.1158/0008-5472.CAN-23-0816

Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C., & Schultz, N. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Science Signaling, 6(269), pl1. https://doi.org/10.1126/scisignal.2004088

Kadoch, C., Hargreaves, D. C., Hodges, C., Elias, L., Ho, L., Ranish, J., & Crabtree, G. R. (2013). Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. Nature Genetics, 45(6), 592–601. https://doi.org/10.1038/ng.2628

Liu, H., Zhao, Y.-R., Chen, B., Ge, Z., & Huang, J.-S. (2019). High expression of SMARCE1 predicts poor prognosis and promotes cell growth and metastasis in gastric cancer. Cancer Management and Research, 11, 3493–3509. https://doi.org/10.2147/CMAR.S195137

Papadakis, A. I., Sun, C., Knijnenburg, T. A., Xue, Y., Grernrum, W., Hölzel, M., Nijkamp, W., Wessels, L. F. A., Beijersbergen, R. L., Bernards, R., & Huang, S. (2015). SMARCE1 suppresses EGFR expression and controls responses to MET and ALK inhibitors in lung cancer. Cell Research, 25(4), 445–458. https://doi.org/10.1038/cr.2015.16

Pries, R., Halske, C., Behrens, H.-M., Krüger, S., & Röcken, C. (2023). SMARCA4 and SMARCE1 in gastric cancer: Correlation with ARID1A, and microsatellite stability, and SMARCE1/ERBB2 co-amplification. Cancer Medicine, 12(8), 9327–9341. https://doi.org/10.1002/cam4.5776

Shain, A. H., & Pollack, J. R. (2013). The spectrum of SWI/SNF mutations, ubiquitous in human cancers. PLoS ONE, 8(1), e55119. https://doi.org/10.1371/journal.pone.0055119

Smith, M. J., O'Sullivan, J., Bhaskar, S. S., Hadfield, K. D., Poke, G., Caird, J., Sharif, S., Eccles, D., Fitzpatrick, D., Rawluk, D., du Plessis, D., Newman, W. G., & Evans, D. G. (2013). Loss-of-function mutations in SMARCE1 cause an inherited disorder of multiple spinal meningiomas. Nature Genetics, 45(3), 295–298. https://doi.org/10.1038/ng.2552

Smith, M. J., Wallace, A. J., Bennett, C., Hasselblatt, M., Elert-Dobkowska, E., Evans, L. T., Hickey, W. F., van Hoff, J., Bauer, D., Lee, A., Hevner, R. F., Beetz, C., du Plessis, D., Kilday, J.-P., Newman, W. G., & Evans, D. G. (2014). Germline SMARCE1 mutations predispose to both spinal and cranial clear cell meningiomas. The Journal of Pathology, 234(4), 436–440. https://doi.org/10.1002/path.4427

Sokol, E. S., Feng, Y.-X., Jin, D. X., Tizabi, M. D., Miller, D. H., Cohen, M. A., Sanduja, S., Reinhardt, F., Pandey, J., Superville, D. A., Jaenisch, R., & Gupta, P. B. (2017). SMARCE1 is required for the invasive progression of in situ cancers. Proceedings of the National Academy of Sciences, 114(16), 4153–4158. https://doi.org/10.1073/pnas.1703931114

St Pierre, R., Collings, C. K., Samé Guerra, D. D., Widmer, C. J., Bolonduro, O., Mashtalir, N., Sankar, A., Liang, Y., Bi, W. L., Gerkes, E. H., Meredith, D. M., & Kadoch, C. (2022). SMARCE1 deficiency generates a targetable mSWI/SNF dependency in clear cell meningioma. Nature Genetics, 54(6), 861–873. https://doi.org/10.1038/s41588-022-01077-0

Tang, Z., Kang, B., Li, C., Chen, T., & Zhang, Z. (2019). GEPIA2: An enhanced web server for large-scale expression profiling and interactive analysis. Nucleic Acids Research, 47(W1), W556–W560. https://doi.org/10.1093/nar/gkz430

Tang, Z., Li, C., Kang, B., Gao, G., Li, C., & Zhang, Z. (2017). GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Research, 45(W1), W98–W102. https://doi.org/10.1093/nar/gkx247