



Qualitative evaluation of MRI features of lipoma and atypical lipomatous tumor: results from a multicenter study

Lorenzo Nardo¹ · Yasser G Abdelhafez¹ · Francesco Acquafredda² · Silvia Schirò^{3,4} · Andrew L. Wong¹ · Dani Sarohia¹ · Roberto Maroldi⁵ · Morgan A. Darrow⁶ · Michele Guindani⁷ · Sonia Lee⁸ · Michelle Zhang⁹ · Ahmed W. Moawad¹⁰ · Khaled M. Elsayes¹¹ · Ramsey D. Badawi¹ · Thomas M. Link³

Received: 12 November 2019 / Revised: 25 December 2019 / Accepted: 1 January 2020 / Published online: 22 January 2020
© ISS 2020

Abstract

Objectives The objectives of the study are (1) to distinguish lipoma (L) from atypical lipomatous tumor (ALT) using MRI qualitative features, (2) to assess the value of contrast enhancement, and (3) to evaluate the reproducibility and confidence level of radiological readings.

Materials and methods Patients with pathologically proven L or ALT, who underwent MRI within 3 months from surgical excision were included in this retrospective multicenter international study. Two radiologists independently reviewed MRI centrally. Impressions were recorded as L or ALT. A third radiologist was consulted for discordant readings. The two radiologists re-read all non-contrast sequences; impression was recorded; then post-contrast images were reviewed and any changes were recorded.

Results A total of 246 patients (135 females; median age, 59 years) were included. ALT was histopathologically confirmed in 70/246 patients. In multivariable analysis, in addition to the lesion size, deep location, proximal lower limb lesions, demonstrating incomplete fat suppression, or increased architectural complexity were the independent predictive features of ALT; but not the contrast enhancement. Post-contrast MRI changed the impression in a total of 5 studies (3 for R1 and 4 for R2; 2 studies are common); all of them were incorrectly changed from Ls to ALTs. Overall, inter-reader kappa agreement was 0.42 (95% CI 0.39–0.56). Discordance between the two readers was statistically significant for both pathologically proven L ($p < 0.001$) and ALT ($p = 0.003$).

Conclusion Most qualitative MR imaging features can help distinguish ALTs from BLs. However, contrast enhancement may be limited and occasionally misleading. Substantial discordance on MRI readings exists between radiologists with a relatively high false positive and negative rates.

Keywords MRI · Adipose tissue neoplasms · Well differentiated liposarcoma · Atypical lipomatous tumor · Lipoma

✉ Lorenzo Nardo
lnardo@ucdavis.edu

¹ Department of Radiology, University of California Davis, 4860 Y Street, Suite 3100, Sacramento, CA 95817, USA

² Radiologia Pediatrica - Ospedale dei Bambini, ASST Spedali Civili, Brescia, Italy

³ Department of Radiology and Biomedical Imaging, University of California, San Francisco, CA, USA

⁴ Section of Radiology, Department of Medicine and Surgery (DiMeC), University of Parma, Parma, Italy

⁵ Scienze Radiologiche, Università degli Studi di Brescia, Brescia, Italy

⁶ Pathology and Laboratory Medicine, University of California Davis, Sacramento, CA, USA

⁷ Department of Statistics, University of California, Irvine, CA, USA

⁸ Department of Radiological Sciences, University of California, Irvine, CA, USA

⁹ Department of Diagnostic Radiology, McGill University Health Center, Montreal, Canada

¹⁰ Department of Diagnostic Imaging, University of Texas MD Anderson Cancer Center, Houston, TX, USA

¹¹ Department of Diagnostic Radiology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Introduction

Lipomatous soft tissue tumors are the most common mesenchymal neoplasms [1]. Lipomatous tumors encompass a wide spectrum ranging from benign to aggressively malignant tumors. While it is not difficult to distinguish high-grade malignant lesions from benign entities, it is more challenging to correctly differentiate lipoma (L) from well differentiated liposarcoma (WDLS) [2]. According to WHO classification, WDLS are generally called atypical lipomatous tumor (ALTs) when located in the extremities or in the trunk to differentiate them from their intrathoracic or intraabdominal counterparts, which are more difficult to completely excise [3].

It is crucial to correctly differentiate L from ALT, because the two tumors undergo different surgical approaches due to their different biological behavior. ALT are generally resected with wide margins as they have a high rate of local recurrence and potential for dedifferentiation into high-grade sarcomas and metastatic spread [4–6]. In addition, ALT need long-term clinical follow-up, as they may exhibit delayed dedifferentiation 5–10 years after resection or recurrence [7, 8]. On the other hand, L can be clinically observed unless they result in symptoms due to the mass effect. In these cases, the mass could be just marginally excised [9].

MRI is the standard modality for assessment of soft tissue neoplasms. Several MR imaging features have been shown to predict the presence of ALT, including tumor size [10, 11], location [10], presence of thick septa [11] or enhancement [12], and aggregated scoring systems which included summing scores from many qualitative features [13, 14]. However, when these features are used to distinguish ALT from BL, there is substantial overlap, which results in low diagnostic accuracy (63%) with specificities as low as 36% [11]. In addition, the radiologic reading reliability has mixed results in the literature, with several previous studies reporting low or fair intra- and inter-observer reproducibility [11, 15]. For these shortcomings, MRI is often deemed not adequate for meeting the clinical needs.

Conventional histopathology relies on the presence of atypical hyperchromatic nuclei as ALT cell hallmark. However, finding these cells is often challenging due to their paucity and their scattered appearance throughout the lesion [16]. For this reason, biopsy cannot be considered appropriate for diagnosis and tumor excision is recommended [17, 18]. Several advanced pathologic tools have been developed to accurately distinguish ALT from L including immunohistochemistry (IHC), fluorescence in situ hybridization (FISH), and molecular testing for antibodies to MDM2 and CDK4. In particular, MDM2 has been described as a highly sensitive marker for ALT; nowadays, it is the most-commonly used technique to

distinguish the most challenging cases of ALT from L [16, 19, 20].

With this multicenter study, we aimed (1) to distinguish L from ALT using MRI qualitative features, (2) to assess the value of contrast enhancement, and (3) to evaluate the reproducibility and confidence level of radiological readings.

Material and methods

This is a retrospective multicenter imaging study. The study protocol was approved by the local Institutional Review Board (IRB No. 1213041-5); the collection/distribution of images was approved and carefully regulated by University of California Reliance System (No. 2963) and Data Transfer Agreements. In addition, the overseas agreements were edited to adhere to the European Union laws in matter of patient data protection.

Study population

The local picture archiving and communications systems of 5 different University Hospitals were queried for the terms: “atypical lipomatous tumor” “soft tissue tumor,” “Lipoma” “soft tissue tumor,” and “MRI.” The search identified 5430 subjects from March 2008 to February 2018.

Subjects were included if they had a pathologically proven surgically resected L or ALT and underwent a preoperative MRI study (with or without contrast) within 3 months before their surgery.

We excluded subjects with incomplete imaging studies, poor quality studies (e.g., motion artifact, suppression issues, metallic artifact), recurrent or persistent tumor, or non-conclusive pathologic report for L or ALT.

After applying inclusion and exclusion criteria, a total of 246 subjects were eligible for this analysis (135 female and 111 male, age range 23 to over 89 years—according to institutional privacy rule all subjects ≥ 90 years old, need to be grouped in a common age category). Medical records were reviewed for patient’s demographics and presence of pain or discomfort (the latter data was not available for two patients).

MRI protocol

MRI studies were performed on 21 different scanners including: Hitachi Airis II, Toshiba Titan, Siemens (Aera, Avanto, Espree, Skyra, Sonata, SymphonyTim, Verio and Trio Tim), Philips (Achieva dStream, Gyroscan intera, Ingenia), and GE (Discovery MR750, MSK extreme, Optima MR450w, Signa excite, Signa genesis, Signa HDe, Signa HDx, Signa HDxt). Scanners operated at either 1.5 or 3.0-Tesla magnetic field strength. The MRI imaging protocol included nonfat-suppressed T1-weighted fast

spinecho (FSE) and T2-weighted fat-suppression FS or STIR sequences for all patients. The sequences included sagittal, coronal, and axial T1 SE and T2 FS FSE as well pre- and post-contrast T1-weighted fat-suppressed images. DICOM images were anonymized and exported for central reading in one of the participating sites. Deidentification procedure was confirmed at the central reading site to adhere to Safe Harbor standards [21].

Radiologic assessment

All images were independently reviewed by two musculoskeletal fellowship-trained radiologists, (reader 1, DS, 1-year experience and reader 2, LN, 3-year experience) for inter-observer reproducibility measurements. In those instances where impressions were not identical, consensus readings were performed with the senior MSK radiologist

(T.M.L., more than 20-year experience); this data sheet was considered the consensus reading. The three radiologists were blinded to any clinical data.

The images were reviewed for lesion site, size, location (superficial or deep to the superficial fascia; i.e., the fascial sheet lying directly beneath the skin [22]), architectural complexity (compared to the surrounding fat, Fig. 1a–c), presence/absence of septa thicker than 2 mm, level of fat suppression (Fig. 1d–i), regular/irregular margins, and presence and pattern of enhancement if any as detailed in Table 1.

In addition to the qualitative features, the two readers recorded three more parameters: (1) the overall impression as either L or ALT, (2) their confidence level about that impression on a 4-point scale ranging from one (least confident) to 4 (most confident), and (3) categorization of their impression on a 5-point diagnostic score that mimics the clinical reporting language [23], ranging from 1 to 5; 1 = consistent with L, 2 =

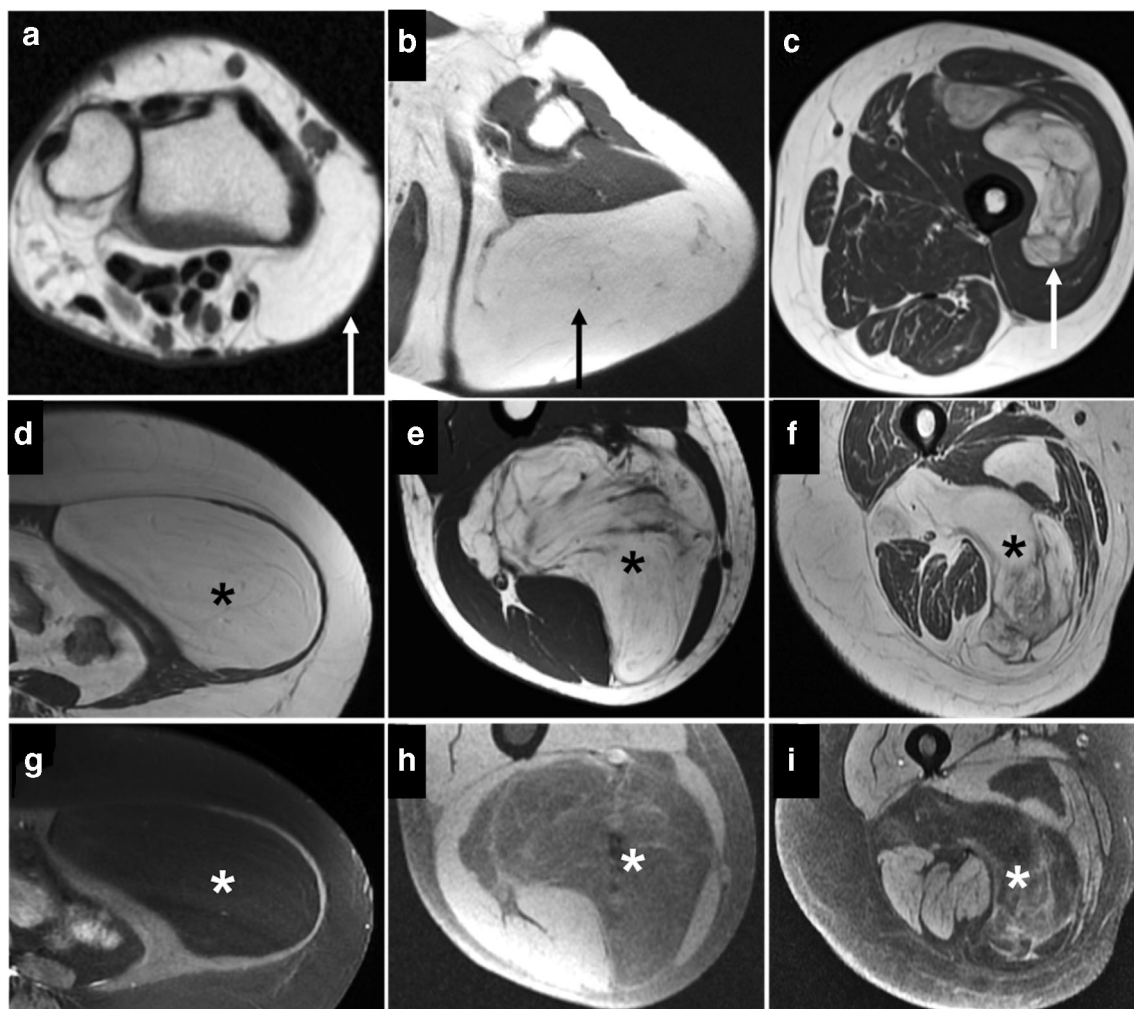


Fig. 1 Levels of lesion architecture complexity (a–c, arrows) in comparison to the surrounding subcutaneous fat are demonstrated in T1W images (top row): **a** less complex architecture, **b** similar complexity, and **c** more complex architecture. T1WI (d, e, f) and the

corresponding T1WI FS (g, h, i) images demonstrating complete suppression of fat signal in (g), near complete suppression in (h), and partial fat suppression in (i)

Table 1 Univariate analysis of the association between the clinical/MR imaging features and the pathological diagnosis of lipomatous lesions

Pathologic diagnosis		Ls		ALTs		OR (95% CI)	p value
		N	%	N	%		
Clinical/MRI features							
Sex	Female	96	54.5%	39	55.7%	0.95 (0.55–1.67)	0.89
	Male	80	45.5%	31	44.3%		
Pain or discomfort [#]	No	81	46.0%	19	27.9%	2.2 (1.2–4.03)	0.01
	Yes	95	54.0%	49	72.1%		
Tumor Site [*]	Proximal UL	53	30.1%	6	8.6%	11.36 (5.72–22.58) ^a	< 0.001
	Distal UL	5	2.8%	1	1.4%		
	Trunk	60	34.1%	4	5.7%		
	Proximal LL	49	27.8%	57	81.4%		
	Distal LL	9	5.1%	2	2.9%		
Tumor Depth [*]	Superficial	71	40.3%	4	5.7%	11.16 (3.89–31.99)	< 0.001
	Deep	105	59.7%	66	94.3%		
Tumor Margin [*]	Smooth	165	93.8%	53	75.7%	4.81 (2.12–10.91)	< 0.001
	Irregular	11	6.3%	17	24.3%		
Fat suppression [*]	Complete	119	67.6%	9	12.9%	14.15 (6.57–30.49) ^b	< 0.001
	Almost complete	38	21.6%	17	24.3%		
	Partial	17	9.7%	42	60.0%		
	Not suppressed	2	1.1%	2	2.9%		
Architecture complexity [*]	Less complex	93	52.8%	11	15.7%	18.2 (9.12–36.34) ^c	< 0.001
	Similar complexity	63	35.8%	10	14.3%		
	More complex	20	11.4%	49	70.0%		
Septal thickness [*]	< 2 mm	152	86.4%	28	40.0%	9.5 (4.99–18.08)	< 0.001
	≥ 2 mm	24	13.6%	42	60.0%		
Enhancement pattern [*]	Non-enhancement	57	39.0%	5	7.7%	7.69 (2.91–20.29) ^d	< 0.001
	Peripheral	10	6.8%	1	1.5%		
	Nodular	3	2.1%	5	7.7%		
	Septal	47	32.2%	21	32.3%		
	Peripheral and nodular	1	0.7%	0	0.0%		
	Nodular and septal	5	3.4%	4	6.2%		
	Peripheral and septal	21	14.4%	22	33.8%		
	Peripheral, nodular and septal	2	1.4%	7	10.8%		

[#] The data were missing in two patients

^{*} These MRI features were based on Reader 2 impressions

^{a,b,c,d} Odds ratio calculated between two groups

^a Proximal lower limb versus other locations

^b Complete versus almost complete, partial, or no suppression

^c More complex versus similar/less complex architecture

^d Presence versus absence of contrast enhancement (post-contrast sequences were available in only 211 patients)

L lipoma, ALT atypical lipomatous tumor, OR odds ratio for the pathologic diagnosis of ALT, CI confidence interval

probably L, 3 = equivocal (possible lipoma, possible ALT), 4 = probably ALT, and 5 = consistent with ALT (Figs. 2, 3 and 4).

The whole dataset was re-reviewed at least 4 weeks from the first reading, to establish intra-observer reliability measurements. During the second reading, the radiologists were asked to record the abovementioned 3 parameters (overall impression, confidence level, and overall diagnostic score) after reviewing pre-contrast MRI sequences only, then

contrast sequences were reviewed and any change in the 3 parameters was recorded.

Pathologic analysis

Pathology reports were reviewed to establish the reference standard for each case. Excision specimens were reviewed microscopically by a pathologist at each site. When the histologic findings were equivocal, a second pathologist was

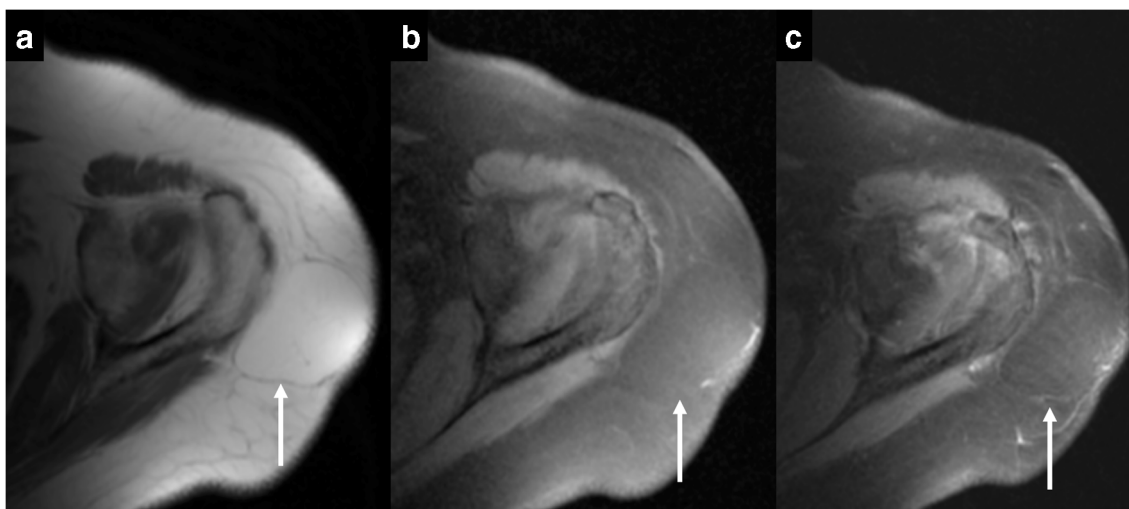


Fig. 2 Example of a lesion with diagnostic score 1-consistent with lipoma (arrow). Shown is a superficial, left shoulder lesion that shows smooth margins, less complex architecture compared to surrounding fat in T1W

images (a) with complete fat suppression in T1 FS image (b) and minimal septal enhancement in T1W post-contrast (c). Conventional pathology was consistent with lipoma

consulted and/or FISH testing was performed to evaluate for the presence of MDM2 gene amplification. This standard assessment reflects the clinical routine in the USA.

Statistical analysis

Association of categorical variables (clinical and MRI features) with the pathologic diagnosis was assessed using Pearson's chi-square or Fisher's exact test, as appropriate. The non-parametric Mann-Whitney *U* test was used for comparing the non-normally distributed continuous variables (age and lesion size) among L and ALT.

Odds ratios were calculated for binary variables. Significant features from the univariate analysis (UVA) were further employed in a multivariable analysis (MVA) using a binary logistic regression model with backward selection method.

Cohen's Kappa test was used for assessing the degree of agreement between any two readers.

True-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN) readings were identified on the basis of subsequent histopathologic validation. Diagnostic performance parameters were calculated in the form of sensitivity, specificity, accuracy, positive predictive value (PPV), negative predictive value (NPV), and likelihood ratios (LRs). The non-parametric McNemar's test was used to evaluate the statistical significance of the differences in sensitivity and specificity between the paired readings. The area under the curve (AUC) from receiver operating characteristic (ROC) analysis was used to demonstrate overall accuracy of each modality.

Continuous data were summarized as mean and standard deviation or median (range). Categorical variables were summarized as frequency and percentages. Confidence intervals were reported where applicable. In all analyses, *p* value < 0.05

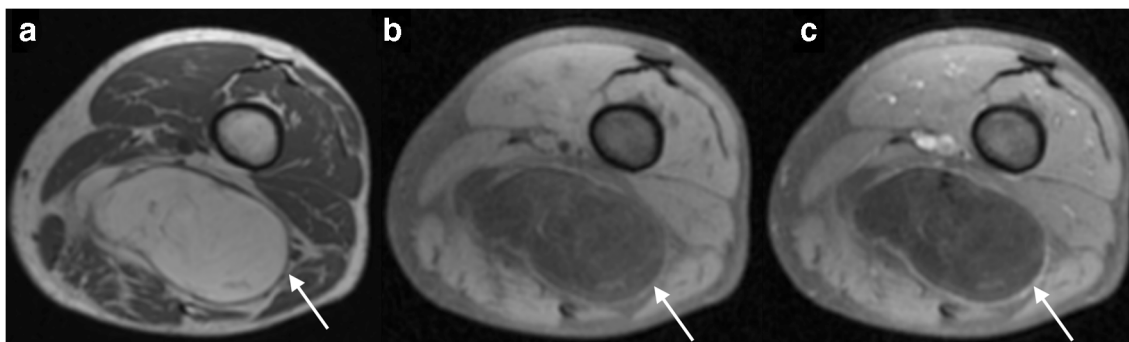


Fig. 3 Example of lesion with diagnostic score 3-equivocal (possible lipoma, possible ALT, arrow). Shown is a posterior thigh lesion with smooth margins and fine septations deep to the fascia. The lesion has architecture similar to surrounding fat in T1W images (a). The lesion

demonstrates incomplete fat suppression in T1 FS image (b) and mild peripheral and septal enhancement in T1W post-contrast sequence (c). Pathology with MDM2 revealed ALT

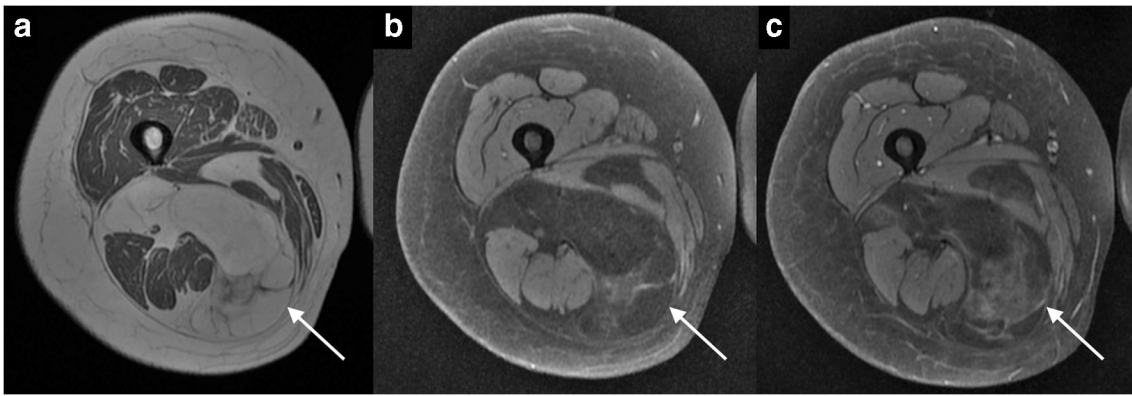


Fig. 4 Example of lesion with diagnostic score 5-consistent with ALT (arrow). Shown is a large posterior thigh lobulated lesion with architecture complexity higher than surrounding fat in T1WI (a). The lesion

demonstrates partial fat suppression in T1 FS image (b) and heterogeneous and nodular enhancement in T1W post-contrast sequence (c) Pathology with MDM2 revealed ALT

(two-tailed) was considered statistically significant. The analysis was performed using SPSS v.21 (IBM Corp, Armonk, New York) and MedCalc (MedCalc Software, Mariakerke, Belgium).

ALT was histopathologically proved in 70/246 patients (prevalence 28%, 95% CI 23–35%). The prevalence of ALT did not vary significantly between the 5 participating sites (range 24–38%).

Results

Study population

A total of 246 subjects (135 females, 111 males; median age 59; range 23 to over 89 years) were eligible for this analysis. The average tumor size was 12.4 ± 7.2 cm. Surgical excision was performed in all subjects and free margins were obtained in 209/246 patients. MDM2 status was assessed in 47% of lesions ($n = 116$). MDM2 was positive in 53 lesions, all of which were ALTs; and negative in 63 lesions, 59 of them were BLs. MDM2 was negative in 4 ALTs ($p < 0.0001$). Overall,

Clinical and MRI features of ALT

Using univariate analyses, subjects with ALT were significantly older (61 ± 13 years) compared to those having lipoma (56 ± 12 years, $p = 0.004$). On average, the maximum lesion size was 18 ± 7 cm for ALTs compared to 10 ± 6 cm for Ls ($p < 0.001$). ALTs were seen more frequently in the proximal lower limb, located deep to the superficial fascia (Fig. 5), and had irregular margins and thick septa. In addition, ALTs often showed contrast enhancement, incomplete fat suppression, and more complex architecture compared to the surrounding fat (Table 1).

Fig. 5 Superficial lipomatous lesions from two different patients: **a** 56-year woman right gluteal mass (arrow), **b** 61-year woman with left gluteal mass. The two masses demonstrate similar MRI features in the shown T1W images: fat architecture similar to the surrounding fat and thin septa are noted. The original read for these lesions was lipoma; however, the histopathologic report was consistent with lipoma in (a), (arrow) and atypical lipomatous tumor in (b), (two arrows)

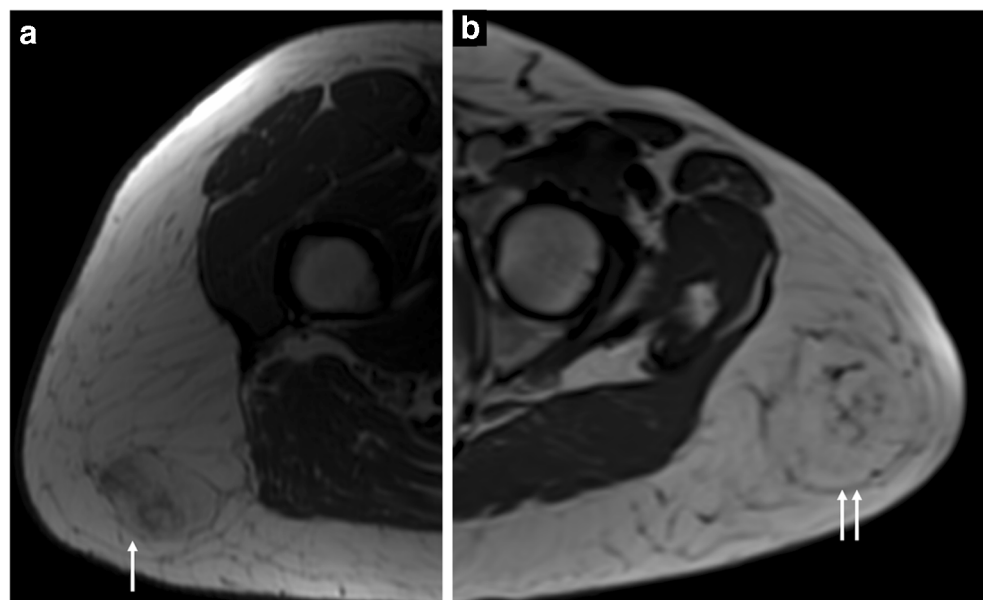


Table 2 Multivariable analysis of the significant qualitative features predictive for the diagnosis of atypical lipomatous tumor

Features	<i>p</i> value	OR (95% CI)
Maximum lesion size (continuous variable)	0.016	1.09 (1.02–1.17)
Site: lower limb	0.002	4.93 (1.81–13.46)
Depth: deep to superficial fascia	0.020	5.92 (1.32–26.62)
Fat suppression: incomplete	0.022	3.43 (1.19–9.89)
Architecture: more complex	< 0.0001	7.80 (2.96–20.59)

OR odds ratio for the pathologic diagnosis of ALT, CI confidence interval

In multivariable analysis, after adjusting for age and sex, the lesion size, proximal lower limb location, deep to superficial fascia, incomplete fat suppression, and increased architectural complexity were independent predictors of ALT (Table 2). However, lesion enhancement was not associated with the diagnosis of ALT.

Impact of contrast injection

Overall, radiological impressions were changed in 7 readings (for a total of 5 studies), all of them were incorrectly changed from L to ALT. Of the 7 incorrect reads, both readers concordantly changed their impression in 2 studies. Additional 1 and 2 studies were reported as positive for ALT after reviewing the contrast sequences from readers 1 and 2, respectively (Fig. 6). Pre-contrast MRI was associated with slightly higher specificity (81%, 95% CI 70–85%) compared to readings that involve post-contrast MRI sequences (75%, 95% CI 68–82%; $p = 0.1$) with identical level of sensitivity (83%, 95% CI 72–91%).

Similarly, the overall accuracy of the two readings on the 5-point scale was comparable with AUCs of 0.89 (95% CI 0.84–0.93) and 0.88 (95% CI 0.83–0.93), for pre- and post-contrast MRI, respectively ($p = 0.1$). Post-contrast MRI did not significantly change the confidence or diagnostic scores.

Agreement on MRI readings

Reader 1 did not report any lesions in one study; a total of 245 scans were available for agreement analysis. The two radiologists agreed on categorizing 90 lesions as L and 87 as ALT. Discordant readings were encountered in 68 studies (28%). Inter-reader kappa agreement was 0.42 (95% CI 0.39–0.56). Discordance between the two readers was statistically significant for both pathologically proven L ($p < 0.001$) and ALT ($p = 0.003$). Among the 68 discordant cases between the first two observers, the third radiologist agreed on 11 of reader 1 and 57 of reader 2 impressions with kappa levels of -0.14 (95% CI -0.29 – 0.02) and 0.39 (95% CI 0.09 – 0.64), respectively. Of the 68 subjects, the false categorization of readers 1, 2, and 3 were 50, 17, and 15, respectively.

Intra-observer reliability of MRI readings was very good with kappa value of 0.97 (95% CI 0.90–1.0) for reader 1 and 0.88 (95% CI 0.82–0.94) for reader 2.

Diagnostic performance of MRI

Consensus MRI readings correctly categorized 62/70 ALTs and 137/176 Ls (Table 3). The false positive and negative rates were 22% and 11%, respectively.

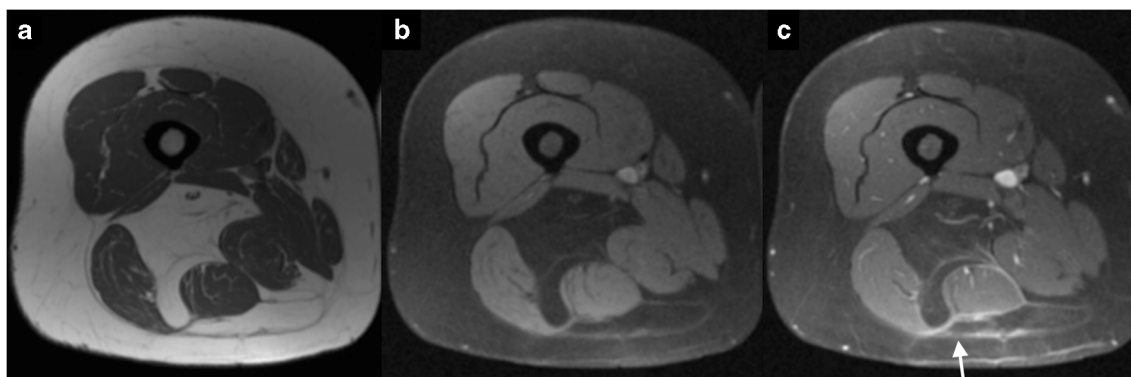


Fig. 6 A 40-year-old woman with a posterior thigh mass. T1WI (a) demonstrates a $9.0 \times 6.1 \times 9.6$ -cm lipomatous lesion between semimembranosus and posterior bundle of vastus lateralis, with architecture similar to subcutaneous fat and thin septa. The lesion demonstrates complete fat suppression in T1 FS pre-contrast image (b). The lesion was

scored 2 (probably lipoma) before reading post-contrast images T1WI (c) which showed significant peripheral enhancement (arrow) and lead the reader to upscale his score to 3 (equivocal). The lesion was surgically excised with free margin, and histopathological assessment with negative MDM2 was consistent with lipoma

Table 3 Diagnostic performance indices for the readings from each radiologist and from the consensus readings

N or value (95% CI)			
Parameter	Reader 1*	Reader 2	Consensus reading
True positive	68	57	62
True negative	95	139	137
False positive	80	37	39
False negative	2	13	8
Sensitivity	97 (90–100%)	81 (70–90%)	89 (79–95%)
Specificity	54 (47–62%)	79 (72–85%)	78 (71–84%)
AUC	0.76 (0.70–0.80)	0.80 (0.75–0.85)	0.83 (0.78–0.88)
+LR	2.12 (1.80–2.51)	3.87 (2.85–5.27)	4.00 (2.99–5.34)
–LR	0.05 (0.01–0.21)	0.24 (0.14–0.39)	0.15 (0.08–0.28)
PPV	46 (38–54)	61 (50–71%)	61 (51–71%)
NPV	98 (93–100)	91 (86–95%)	94 (89–98%)

AUC area under the receiver operating characteristics curve, +LR positive likelihood ratio, –LR negative likelihood ratio, PPV positive predictive value, NPV negative predictive value

*No lesion was reported in one study by reader 1 and read as lipoma by the other two readers

Radiologic confidence

The percentage of incorrect MRI impressions correlated negatively with the confidence score (Table 4). Within the readings that were given the lowest confidence score by readers 1 and 2, respectively, 80% and 40% of their final impressions were incorrect compared to only 19% and 8% for the highest confidence score.

Discussion

In this multicenter study, we demonstrated that several MRI features can help differentiating benign from malignant lipomatous lesions; however, the clinical reading suffers from relatively low level of confidence and reproducibility even in the hands of experienced radiologists.

Our study demonstrated that several clinical and MRI features are independently predictive of ALTs, including lesion

size, proximal lower limb location, deep site, incomplete fat suppression, and increased architectural complexity. These findings confirm the results of prior studies [10, 11, 13, 14]. However, in contrast to other studies [12, 24], thick septa and enhancement were associated with the diagnosis of ALT only in univariate analysis.

We have tried to assess the value of use of intravenous contrast on the reading accuracy. Previous studies have demonstrated that the presence of gadolinium enhancement is predictive of ALT [12, 24]. To our surprise, contrast enhancement was only significantly associated with ALT in univariate analysis. Sequential reading of MRI without, then with, the addition of contrast sequences changed the radiologist's impression in a total of only 5 cases from the two readers; all of them were pathologically proven lipomas that showed different levels/patterns of gadolinium enhancement and lead to the false impression of ALT diagnosis. The specificity of post-contrast MRI was 75% (95% CI 68–82%) which is comparable to prior studies [17]. Our large multicenter study suggests that the value of contrast administration may be far limited than previously reported [12–14, 24]. Acquisition of contrast-enhanced MRI sequences increases the time and cost related to the procedure. Also, gadolinium administration has been associated with possible side effects including nephrogenic systemic fibrosis and accumulation in the basal ganglia [25–31]; therefore, we believe that careful selection of the population that may benefit the most from contrast administration after obtaining pre-contrast MRI images would decrease the potential risks associated with gadolinium injection and reduce the overall scan time and, accordingly, the overall study cost. We understand that our retrospective findings, from a well-selected dataset cannot represent the basis for changing the employed clinical protocols; however, it is a step towards personalized radiological approach.

The reproducibility of MRI readings in our study was moderate with inter-rater agreement of 0.42. Previous reports demonstrated inter-observer agreement on the final radiological impression that ranged from 0.63 [15] to 0.71 [11], with the agreement on each MRI feature was highly variable, reaching as low as 0.17 [15]. Our figures are on the lower side of these values. The fact that our large cohort came from 5 different institutions, each with different MRI acquisition/processing

Table 4 Number and percentage of correct and incorrect impressions from the two radiologists according to the confidence score

Observers Radiologic confidence score	Reader 1*		Reader 2	
	Incorrect N (%)	Correct N (%)	Incorrect N (%)	Correct N (%)
Score 1	12 (80%)	3 (20%)	14 (40%)	21 (60%)
Score 2	30 (83%)	6 (17%)	19 (37%)	32 (63%)
Score 3	13 (24%)	42 (76%)	9 (14%)	54 (86%)
Score 4	27 (19%)	112 (81%)	8 (8%)	89 (92%)

*No lesion was reported in one study by reader 1 and read as lipoma by the other two readers

protocols, may, at least partially, explain these findings. Also, the spectrum of findings seen in our large cohort reflected multiple exceptions to the known common rules for the diagnosis of ALT. For example, we encountered few ALTs that were superficially located ($n = 4$), smaller than 10 cm ($n = 13$) or even smaller than 5 cm ($n = 1$). These features could mislead the radiologist(s) and contribute to the false negative results encountered.

Our study has several limitations. First, since this is a retrospective multicenter study, substantial heterogeneity exists regarding the type of MRI machines, acquisition and processing methodology, and selection of subjects for different clinical algorithms including surgical intervention. The central reading by two experienced radiologists may have addressed, at least partially, this concern. Second, the assessment of the value of contrast injection was performed only in 86% of the study population since 35 subjects did not have contrast sequences. However, this was a pre-made decision as most institutions are pre-protocolling their patients' procedure before the patient is actually being scanned. Therefore, this data reliably represents routine care. Third, some studies showed artifacts from surface coil field inhomogeneity which could limit the evaluation of the fat suppression sequence, especially for the most superficial lesions; to minimize this concern, all efforts have been made to exclude all images with poor quality. Fourth, this is a cross-sectional study, with no imaging follow up. Since several surgeons advocate for wait-and-watch decisions, not only for lipomas but also for ALTs, a longitudinal study would help us understand the evolution of imaging features over time, and also estimate the recurrence of resected lipomatous lesions. Fifth, MDM2 was not performed in all our patients as the pathologic reference test. In clinical practice, MDM2 may not be requested except when the conventional histopathology is equivocal; therefore, even though we recognize this limitation, this data can be considered a real snapshot of the normal clinical routine of academic institutions. Sixth, in clinical practice, a predominantly lipomatous mass could encompass a wide range of heterogeneous differential diagnoses, that fall largely beyond the two options presented to the radiologists in the current study; accordingly, the results presented here may not reflect the true diagnostic yield when applied in the clinical setting. Also, since we included only lesions that were surgically resected and proved to be one of the two diagnoses, our results may be biased by the large proportion of predominantly fat containing lesions that are confidently diagnosed by MRI and did not undergo any resection. Seventh, we did not systemically study the finding of fat necrosis which could be challenging to differentiate from liposarcoma on imaging and needs histopathologic evaluation [1]. Last, given the logistic and administrative restrictions on data sharing, all the data and images collected from the 5 institutions were completely anonymized. It was not possible to re-check additional patients' data.

The advantages of the current study include its multicenter design which allowed the collection of, to the best of our knowledge, the largest MRI database of pathologically confirmed ALTs and BLs. Also, the central reading by 2 experienced musculoskeletal radiologists and consulting a third reader from a different institution for discordant cases rather than solving them between the two radiologists, may decrease the additional bias that could otherwise have resulted from the differences in training and experience between radiologists. Furthermore, we adopted a 5-point diagnostic score that mimics the language used in daily radiologic reporting [23] and used at top cancer specialized institutions such as the Memorial Sloan Kettering Cancer Center.

In conclusion, qualitative MRI features can help distinguish ALTs from BLs; however, a significant overlap may exist between the two conditions. The added value of contrast enhancement may be limited and may not improve the confidence of the radiologic reading. Further work is warranted to optimize a clinical algorithm for selecting the patients undergoing gadolinium administration, possibly including precontrast imaging assessment in the algorithm. Furthermore, substantial discordance on MRI readings exists between well-trained radiologists. Additional work, implementing artificial intelligence/machine learning approaches, is being sought to further analyze MRI images, extract radiomic features and explore the potential impact on the reproducibility, confidence and overall accuracy of the qualitative reading.

Funding information This study is supported in part by the UC Davis Paul Calabresi Career Development Award for Clinical Oncology as funded by the National Cancer Institute/National Institutes of Health through grant no. 5K12-CA138464.

Compliance with ethical standards IRB#: 1213041-5

Conflict of interest Ramsey D Badawi received research/grant support from United Imaging Healthcare.

The authors declare that they have no conflict of interest.

References

1. Gupta P, Potti TA, Wuertzer SD, Lenchik L, Pacholke DA. Spectrum of fat-containing soft-tissue masses at MR imaging: the common, the uncommon, the characteristic, and the sometimes confusing. *Radiographics*. 2016;36(3):753–66.
2. Evans HL, Soule EH, Winkelmann RK. Atypical lipoma, atypical intramuscular lipoma, and well differentiated retroperitoneal liposarcoma: a reappraisal of 30 cases formerly classified as well differentiated liposarcoma. *Cancer*. 1979;43(2):574–84.
3. Jo VY, Fletcher CD. WHO classification of soft tissue tumours: an update based on the 2013 (4th) edition. *Pathology*. 2014;46(2):95–104.
4. Lucas DR, Nascimento AG, Sanjay BK, Rock MG. Well-differentiated liposarcoma. The Mayo Clinic experience with 58 cases. *Am J Clin Pathol*. 1994;102(5):677–83.

5. Weiss SW, Rao VK. Well-differentiated liposarcoma (atypical lipoma) of deep soft tissue of the extremities, retroperitoneum, and miscellaneous sites. A follow-up study of 92 cases with analysis of the incidence of “dedifferentiation”. *Am J Surg Pathol*. 1992;16(11):1051–8.
6. Rozental TD, Khoury LD, Donthineni-Rao R, Lackman RD. Atypical lipomatous masses of the extremities: outcome of surgical treatment. *Clin Orthop Relat Res*. 2002;398:203–11.
7. Gaskin CM, Helms CA. Lipomas, lipoma variants, and well-differentiated liposarcomas (atypical lipomas): results of MRI evaluations of 126 consecutive fatty masses. *AJR Am J Roentgenol*. 2004;182(3):733–9.
8. Mavrogenis AF, Lesensky J, Romagnoli C, Alberghini M, Letson GD, Ruggieri P. Atypical lipomatous tumors/well-differentiated liposarcomas: clinical outcome of 67 patients. *Orthopedics*. 2011;34(12):e893–8.
9. Laurino L, Furlanetto A, Orvieto E, Dei Tos AP. Well-differentiated liposarcoma (atypical lipomatous tumors). *Semin Diagn Pathol*. 2001;18(4):258–62.
10. Kransdorf MJ, Bancroft LW, Peterson JJ, Murphey MD, Foster WC, Temple HT. Imaging of fatty tumors: distinction of lipoma and well-differentiated liposarcoma. *Radiology*. 2002;224(1):99–104.
11. Brisson M, Kashima T, Delaney D, et al. MRI characteristics of lipoma and atypical lipomatous tumor/well-differentiated liposarcoma: retrospective comparison with histology and MDM2 gene amplification. *Skelet Radiol*. 2013;42(5):635–47.
12. Knebel C, Neumann J, Schwaiger BJ, et al. Differentiating atypical lipomatous tumors from lipomas with magnetic resonance imaging: a comparison with MDM2 gene amplification status. *BMC Cancer*. 2019;19(1):309.
13. Cheng Y, Ko AT, Huang JH, et al. Developing a clinical scoring system to differentiate deep-seated atypical lipomatous tumor from lipoma of soft tissue. *Asian J Surg*. 2019.
14. Nagano S, Yokouchi M, Setoguchi T, et al. Differentiation of lipoma and atypical lipomatous tumor by a scoring system: implication of increased vascularity on pathogenesis of liposarcoma. *BMC Musculoskelet Disord*. 2015;16:36.
15. O'Donnell PW, Griffin AM, Eward WC, et al. Can experienced observers differentiate between lipoma and well-differentiated liposarcoma using only MRI? *Sarcoma*. 2013; 2013:982784.
16. Zhang H, Erickson-Johnson M, Wang X, et al. Molecular testing for lipomatous tumors: critical analysis and test recommendations based on the analysis of 405 extremity-based tumors. *Am J Surg Pathol*. 2010;34(9):1304–11.
17. Panzarella MJ, Naqvi AH, Cohen HE, Damron TA. Predictive value of gadolinium enhancement in differentiating ALT/WD liposarcomas from benign fatty tumors. *Skelet Radiol*. 2005;34(5):272–8.
18. Ikoma N, Torres KE, Somaiah N, et al. Accuracy of preoperative percutaneous biopsy for the diagnosis of retroperitoneal liposarcoma subtypes. *Ann Surg Oncol*. 2015;22(4):1068–72.
19. Dei Tos AP, Doglioni C, Piccinin S, et al. Coordinated expression and amplification of the MDM2, CDK4, and HMGI-C genes in atypical lipomatous tumours. *J Pathol*. 2000;190(5):531–6.
20. Kashima T, Halai D, Ye H, et al. Sensitivity of MDM2 amplification and unexpected multiple faint alpha 12 (alpha 12 satellite sequences) signals in atypical lipomatous tumor. *Mod Pathol*. 2012;25(10):1384–96.
21. U.S. Department of Health & Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the Health Insurance Portability and Accountability Act (HIPAA) privacy rule. 2019.
22. Lee S, Joo KB, Song SY. Accurate definition of superficial and deep fascia. *Radiology*. 2011;261(3):994.
23. Panicek DM, Hricak H. How sure are you, doctor? A standardized lexicon to describe the radiologist's level of certainty. *AJR Am J Roentgenol*. 2016;207(1):2–3.
24. Ohguri T, Aoki T, Hisaoka M, et al. Differential diagnosis of benign peripheral lipoma from well-differentiated liposarcoma on MR imaging: is comparison of margins and internal characteristics useful? *AJR Am J Roentgenol*. 2003;180(6):1689–94.
25. Agarwal R, Brunelli SM, Williams K, Mitchell MD, Feldman HI, Umscheid CA. Gadolinium-based contrast agents and nephrogenic systemic fibrosis: a systematic review and meta-analysis. *Nephrol Dial Transplant*. 2009;24(3):856–63.
26. Kanda T, Fukusato T, Matsuda M, et al. Gadolinium-based contrast agent accumulates in the brain even in subjects without severe renal dysfunction: evaluation of autopsy brain specimens with inductively coupled plasma mass spectroscopy. *Radiology*. 2015;276(1):228–32.
27. McDonald RJ, McDonald JS, Kallmes DF, et al. Gadolinium deposition in human brain tissues after contrast-enhanced MR imaging in adult patients without intracranial abnormalities. *Radiology*. 2017;285(2):546–54.
28. Rogosnitzky M, Branch S. Gadolinium-based contrast agent toxicity: a review of known and proposed mechanisms. *BioMetals*. 2016;29(3):365–76.
29. McDonald RJ, McDonald JS, Kallmes DF, et al. Intracranial gadolinium deposition after contrast-enhanced MR imaging. *Radiology*. 2015;275(3):772–82.
30. Idée J-M, Fretellier N, Robic C, Corot C. The role of gadolinium chelates in the mechanism of nephrogenic systemic fibrosis: a critical update. *Crit Rev Toxicol*. 2014;44(10):895–913.
31. Behzadi AH, Zhao Y, Farooq Z, Prince MR. Immediate allergic reactions to gadolinium-based contrast agents: a systematic review and meta-analysis. *Radiology*. 2018;286(2):471–82.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.