doi.org/10.36721/PJPS/2026.40.17148.1

Baseline high-frequency ultrasound thickness and doppler vascularity predict early imaging response to intralesional corticosteroid therapy in keloids: A compound betamethasone cohort study

Tianyi Xie and Xiaodong Chen*
Department of Dermatology, Affiliated Hospital of Nantong University, Medical School of Nantong University, Nantong, Jiangsu, 226001, China

Abstract: **Background**: High-frequency ultrasound (HFUS) parameters such as scar thickness and intralesional vascularity may reflect keloid biological activity and predict treatment response, yet evidence supporting their baseline predictive value remains limited. **Objectives**: To evaluate whether baseline HFUS-derived thickness and Doppler vascularity can predict early imaging response to intralesional corticosteroid therapy in keloids and to explore their association with fibrosis-related molecular signatures. **Methods**: This single-center prospective cohort study enrolled 150 keloid patients receiving standardized compound betamethasone injections. Baseline HFUS measurements included scar thickness, Adler vascularity grade and shear wave elastography (SWE)-derived elastic modulus, along with Vancouver Scar Scale (VSS) assessment. We defined early imaging response at week 4 as percentage thickness reduction ($\Delta Th\%_4W$). Multivariable logistic regression and receiver operating characteristic (ROC) analysis identified predictors of poor early response ($\Delta Th\%_4W < 30\%$). We analyzed an exploratory molecular subgroup (n=40) for miR-205-5p and RUNX2 expression. **Results**: Baseline thickness (OR=2.41) and Doppler vascularity independently predicted poor early response. The combined HFUS model showed good discriminative performance (AUC=0.90). In the exploratory molecular subgroup, Doppler-positive lesions showed lower miR-205-5p and higher RUNX2 expression; RUNX2 correlated positively with thickness (r=0.92). At the week-24 follow-up (n=136), the overall recurrence rate was 20.6%, with higher rates in high-activity lesions (52.6%) than in low-activity lesions (7.1%). **Conclusion**: Baseline HFUS thickness and Doppler vascularity may be potential non-invasive predictors of early response to intralesional corticosteroid therapy in keloids. The HFUS-based activity classification shows clinical potential in identifying patients likely to respond poorly to injection monotherapy, but requires external validation. The observed molecular associations are exploratory and need further functional validation.

Keywords: Early imaging response; High-frequency ultrasound; Intralesional vascularity; Keloid scar; Personalized medicine

Submitted 17-07-2026 – Revised 24-09-2026 – Accepted 30-09-2026

INTRODUCTION

Keloids have a high tendency to recur after clinical intervention. Epidemiological data indicate that recurrence rates after simple surgical excision range from 67% to 100%; even with adjuvant therapies such as local glucocorticoid injections, the overall recurrence rate remains alarmingly high at 54%–73% (Merlino *et al*., 2025). This reality suggests that current treatment protocols relying primarily on empirical injection lack a basis for individualized stratification—some patients show poor efficacy within weeks of treatment initiation and delayed identification leads to unnecessary drug exposure and economic burden. Although novel delivery systems, such as 5-fluorouracil dissolving microneedle patches, have achieved an average thickness reduction of 2.85 mm after 12 weeks, approximately 35% of lesions remain treatment-resistant (Leelawattanachai *et al*., 2025). These findings highlight the urgent need for reliable biomarkers that can accurately predict efficacy before or early in treatment.

Although local glucocorticoid injection is a first-line treatment, clinical efficacy evaluation has long relied on subjective scales (such as the Vancouver Scar Scale), visual observation and clinical experience. Keloid scars show significant heterogeneity across individuals and lesion stages, and traditional methods fail to capture dynamic changes in internal tissue structure, making it difficult to identify poor responders early in the treatment course. In recent years, high-frequency ultrasound (HFUS) has emerged as a pivotal tool for assessing fibroproliferative skin diseases because of its non-invasiveness, real-time capability, high resolution, and reproducibility (Wen *et al*., 2026). Longitudinal monitoring via high-frequency ultrasound has shown that intralesional corticosteroid therapy can produce a mean reduction in keloid dermal thickness of approximately 39.0% (Cai *et al*., 2020), suggesting that ultrasound-derived thickness change can serve as an objective indicator of treatment response. However, most existing studies have focused on dynamic monitoring during treatment; whether baseline parameters can predict subsequent efficacy before the first injection remains unknown. From a clinical decision-making

**Corresponding author:* e-mail: dermatochen@126.com

Article in Press

perspective, predicting treatment responsiveness before injection would enable proactive individualized planning and avoid the time and economic costs of a trial-and-error approach.

Among the various HFUS parameters, scar thickness and intralesional vascularity possess the most direct biological plausibility. Increased thickness prolongs the drug diffusion pathway, while vascularity reflects the lesion's inflammatory activity and perfusion status—abundant blood flow may accelerate drug clearance and sustain fibroblast activation (Willenborg *et al.,* 2025; Cho *et al*., 2024). Therefore, baseline thickness and vascularity are rational candidate biomarkers for predicting injection treatment response.

A second innovative dimension of this study lies in correlating ultrasound-defined imaging phenotypes with fibrosis-related molecular markers. Previous imaging-based studies have largely remained at the parameter-description level without exploring the molecular drivers underlying imaging features. miR-205-5p is a critical negative regulator of the TGF-β signaling pathway that limits excessive collagen gene transcription (Prananda *et al*., 2025; Kong *et al*., 2025); RUNX2 activates collagen gene transcription and synergizes with Smad3 to enhance the pro-fibrotic effects of TGF-β signaling (Hamzehlou *et al*., 2026; Tang *et al*., 2025a). These two molecules may constitute a negative regulatory axis (Zhang *et al*., 2020; Mo *et al*., 2021). However, whether these molecular markers are associated with the ultrasound-defined "high-activity" imaging phenotype has not been reported. Therefore, the association between ultrasound parameters and the miR-205-5p/RUNX2 expression profile was explored to provide exploratory biological validation for imaging-based stratification.

In summary, distinct from previous studies focusing on intra-treatment monitoring or single-parameter description, this study offers two levels of innovation: (1) forward-shifting the predictive window—evaluating whether baseline HFUS thickness and vascularity independently predict Week-4 early imaging response, extending HFUS application from "intra-treatment monitoring" to "pre-treatment prediction"; and (2) imaging-molecular integration—exploring whether ultrasound-based activity stratification is associated with the miR-205-5p/RUNX2 expression profile, offering preliminary biological plausibility for ultrasound-based typing.

This study is in the early development phase of prediction modeling, using prospectively collected data to build a predictive model and evaluate its preliminary performance through exploratory analysis, rather than a pre-registered confirmatory study of predefined predictors. The molecular analyses are explicitly exploratory, based on a small non-randomized subgroup and their findings require functional validation. Readers should distinguish between the primary predictive findings (which have undergone internal validation) and the exploratory molecular associations (which are hypothesis-generating only). Therefore, the proposed ultrasound-based stratification system requires external validation in independent cohorts before clinical implementation can be considered.

The present work has two distinct components with different evidentiary standards. First, the primary predictive analysis evaluates whether baseline HFUS thickness and intralesional vascularity predict early imaging response (Week 4) to intralesional corticosteroid therapy in keloids, using multivariable logistic regression with internal validation. Second, in a separate exploratory molecular substudy, whether ultrasound-defined imaging phenotypes are associated with miR-205-5p and RUNX2 expression was investigated in a small subgroup (n = 40), with the aim of generating hypotheses for future mechanistic studies rather than establishing definitive biological links. These two components are analytically independent and should be interpreted with different levels of confidence.

MATERIALS AND METHODS

Study design and setting

This study was a single-center, prospective cohort study with exploratory model development, conducted at Affiliated Hospital of Nantong University between January 2023 and December 2025. The study followed the Declaration of Helsinki and reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines.

All eligible patients were enrolled consecutively from outpatient clinics without selective sampling. Two independent research coordinators assessed each patient's eligibility according to the pre-specified inclusion and exclusion criteria to minimize selection bias. All baseline assessments, therapeutic interventions and follow-up visits were completed within the predefined study time windows.

Study population

Inclusion and exclusion criteria

A total of 150 keloid patients who received intralesional injection therapy at the study hospital between 2023 and 2025 were enrolled.

Inclusion criteria: (1) Clinical diagnosis confirmed as keloid according to the diagnostic criteria of the International Clinical Recommendations on Scar Management (Gold *et al*., 2014), requiring all of the following: scar extending beyond the original wound boundaries (crab-like or claw-like growth pattern); duration ≥6 months without spontaneous regression; accompanied by pruritus and/or pain; no prior treatment of

In Press

the target lesion within the previous 6 months. In cases where clinical diagnosis was equivocal, punch biopsy (3–4 mm) was performed for histopathological confirmation. (2) Lesions were considered localized (longest diameter ≤ 10 cm; no sonographic evidence of subcutaneous fat invasion). (3) Age 18–65 years. (4) Lesion location confined to the trunk or extremities (excluding genitalia, perineum and mucosal surfaces). (5) Complete baseline clinical and ultrasound documentation available. (6) No prior treatment of the target lesion as specified in the exclusion criteria below.

Exclusion criteria: (1) Prior treatment of the target lesion: intralesional injection within 6 months; laser therapy within 3 months; prior surgical excision at any time; radiotherapy within 12 months; systemic antifibrotic/immunosuppressive therapy within 4 weeks. (2) Systemic contraindications to intralesional corticosteroid therapy: uncontrolled hypertension (systolic >160 mmHg or diastolic >100 mmHg); uncontrolled diabetes mellitus (HbA1c >8.0%); active peptic ulcer disease; severe osteoporosis; active tuberculosis or systemic fungal infection. (3) Organ dysfunction: significant hepatic impairment (ALT/AST >3× upper limit of normal) or renal impairment (eGFR <30 mL/min/1.73m²). (4) Coagulation disorders (platelet count $<50 \times 10^9$/L; INR >1.5; use of anticoagulants not withdrawable for injection). (5) Confirmed malignancy (any type, within the past 5 years). (6) Psychiatric disorders or cognitive impairment that would preclude informed consent or adherence to follow-up. (7) Known hypersensitivity to betamethasone or any excipient in the injection formulation. (8) Pregnancy or lactation (confirmed by urine or serum β-hCG test within 72 hours prior to the first injection). (9) Severe lesions with extreme tension abnormalities or significant joint functional impairment requiring surgical reconstruction as first-line treatment, as determined by the treating physician.

Supplementary note on exclusion criterion (9): This criterion was designed to exclude only the most severe subgroup of lesions for which intralesional corticosteroid monotherapy would be clinically inappropriate as initial therapy. Specifically, it applies to: (1) lesions with extreme tension abnormalities that would fundamentally compromise drug distribution and efficacy, or (2) lesions causing significant joint functional impairment where surgical reconstruction is the preferred first-line approach. It does not exclude anterior chest lesions or joint-region lesions per se. Keloids located on the anterior chest or over joint surfaces that do not meet these severity thresholds—and for which intralesional injection remains a reasonable first-line option—were eligible for enrollment. This distinction is reflected in the representative ultrasound images, which include anterior chest lesions (Figs. 1, 2C, 2D and 3) and a knee lesion (Fig. 2B). The rationale for excluding only the severe subgroup is that: (1) extreme tension abnormalities may independently affect drug distribution and efficacy; (2) lesions with significant joint functional impairment often present with treatment goals that differ fundamentally from those of non-functional sites; and (3) such lesions typically require combined surgical or radiotherapeutic intervention rather than injection monotherapy. All enrolled lesions, including those on the anterior chest and joint regions, were confirmed to have no extreme tension abnormalities or significant joint functional impairment requiring surgical reconstruction as first-line treatment.

Data collection

Baseline clinical data collected included age, disease duration, sex, lesion location, multiplicity, etiology, family history and past history. Lesion location was classified into two main categories: truncal and extremity. Etiology was categorized as infection-related or trauma-related. Lesion size was recorded as the longest diameter (cm) measured by ultrasound.

Selection of index lesion in patients with multiple keloids

For patients with more than one qualifying lesion, a single index lesion was selected according to the following hierarchical criteria: (1) The largest lesion (by longest diameter); (2) If diameters were equivalent, the lesion with the highest baseline Adler vascularity grade; (3) If vascularity grade was equivalent, the lesion with the greatest baseline thickness measured by HFUS; (4) If all of the above were equivalent, the lesion located at the site with the highest mechanical tension. All analyses were based on the selected index lesion. This selection strategy may not fully represent the overall disease burden in patients with multiple lesions, and treatment response in one lesion may not reflect the natural history of untreated lesions—a design limitation of this study.

Sample size calculation

The primary outcome was the thickness change ratio at 4 weeks of treatment ($\Delta Th\%_4W$). The mean difference between the two groups was prespecified as 15% (mean of 40% in the favorable early-response group and 25% in the poor early-response group), with a pooled standard deviation of 12%; the significance level was set at $\alpha = 0.05$ (two-tailed) and the statistical power (1–β) was set at 0.80. The sample size was calculated using the formula for two independent-sample t-tests: $n = ((Z\alpha/2 + Z\beta)^2 \times 2 \times \sigma^2) / \delta^2$, where $Z\alpha/2 = 1.96$, $Z\beta = 0.84$, $\sigma = 12\%$, and $\delta = 15\%$. Substituting these values yielded $n = (1.96 + 0.84)^2 \times 2 \times 12^2 / 15^2 = 7.84 \times 288 / 225 \approx 10.0$, indicating that approximately 10 cases were required per group. Given that prediction model development typically requires a total sample size of 10 to 20 times the number of independent variables, and accounting for potential imbalance between groups, a total of 120 cases (approximately 60 per group) was determined. Further

accounting for an anticipated 20% attrition rate, 150 patients were planned for enrollment.

The sample size calculation was based on a two-group comparison of the primary outcome ($\Delta Th\%_4W$) rather than strictly on prediction model development requirements. This approach is commonly adopted in the preliminary phase of exploratory prediction research; however, it should be acknowledged that the predictive model findings from this study are exploratory and require external validation in an independent prospective cohort before generalization.

Ultrasound examination

Equipment and operator standardization

All ultrasound examinations were performed using a high-frequency ultrasound system equipped with a 20-MHz linear-array transducer (Vevo MD, FUJIFILM VisualSonics, Toronto, Canada). Two independent sonographers with over five years of specialized experience in high-frequency skin ultrasound performed the examinations. The two sonographers performed measurements independently without discussion and each measured all parameters in triplicate and averaged the values. Both sonographers were blinded to patients' efficacy grouping and were not involved in clinical treatment decisions.

Before formal enrollment, the two sonographers completed a 2-week standardized training and calibration program covering image acquisition protocols, anatomical landmark identification, measurement plane selection, and parameter interpretation. After training, a pilot test was conducted on 10 non-enrolled keloid patients to calculate the inter-observer intraclass correlation coefficient (ICC); the formal study was initiated only when all key parameters achieved an ICC $\geqslant$ 0.85. Quality control calibrations were performed every 3 months during the study. To assess reliability, each sonographer measured the same lesion in 20 randomly selected non-enrolled patients at two time points ≥24 hours apart to evaluate intra-observer reliability; the two sonographers independently measured the same set of patients (n=20) to evaluate inter-observer reliability. Table 2 shows ICC results for all parameters, with all intra-observer and inter-observer ICC values ≥0.90.

Image acquisition and parameter measurement

Patient positioning: Patients were positioned supine or seated (truncal lesions supine; extremity lesions seated), with the affected area fully exposed and surrounding muscles relaxed. Standardized positional landmarks (e.g., midclavicular line, anterior axillary line) were used to record patient position and lesion orientation, ensuring reproducible positioning at follow-up visits.

Imaging settings: Gain 40–60 dB, imaging depth 5–15 mm (adjusted according to lesion thickness), focal zone placed at the center of the lesion. Two-dimensional grayscale ultrasound was performed in single-focus mode at the highest resolution setting (20 MHz). Color Doppler flow imaging was performed in micro-flow mode with wall filter set to "low" and velocity scale adjusted to the lowest detectable range (1.5–3.0 cm/s) to maximize visualization of low-velocity blood flow. All settings were saved as a system preset and identical parameter presets were used for subsequent follow-up visits of the same patient.

Coupling and probe handling: After applying an adequate amount of ultrasound gel to the lesion surface, a standardized acoustic stand-off pad (approximately 3 mm thickness, AQUAFLEX, Model ZT-3, Beijing Zhongke Medical) was placed over the gel and air bubbles between the skin and pad were expelled. The transducer was lightly placed on the pad surface, with just enough contact to ensure coupling without compressing the pad, avoiding direct pressure on the lesion tissue. The transducer was held perpendicular (90° ± 5°) to the skin surface, with constant contact pressure maintained throughout the scan.

Scanning and plane selection: The lesion was centered within the field of view, with depth and gain adjusted to visualize the entire lesion and at least 1 cm of adjacent healthy skin (representative grayscale ultrasound images are shown in Fig.1). The short-axis measurement plane was defined as the cross-section perpendicular to the lesion's longest axis, passing through the point of maximum elevation; the long-axis plane was defined as the longitudinal section along the longest axis, passing through the maximum diameter.

Measured parameters: Maximum scar thickness: Vertical distance from the epidermal surface at the highest point of the scar to the dermal-subcutaneous junction on the short-axis section. Scar short-axis length: Maximum horizontal diameter on the short-axis section. Supralesional height: Vertical distance from the highest point of the scar to the line connecting the epidermal surfaces of adjacent normal skin. Echo distribution and signal intensity: Echo distribution was classified as homogeneous or heterogeneous; signal intensity was classified as hypoechoic, isoechoic, or hyperechoic. Adler Doppler vascularity grade (representative color micro-flow Doppler images and Adler grading criteria are shown in figure 2 grade 0 (no signal), grade I (1–2 punctate or thin linear signals), grade II (3–4 punctate signals or one obvious vessel), grade III (≥5 punctate vessels or ≥2 longer vessels).

Article in Press

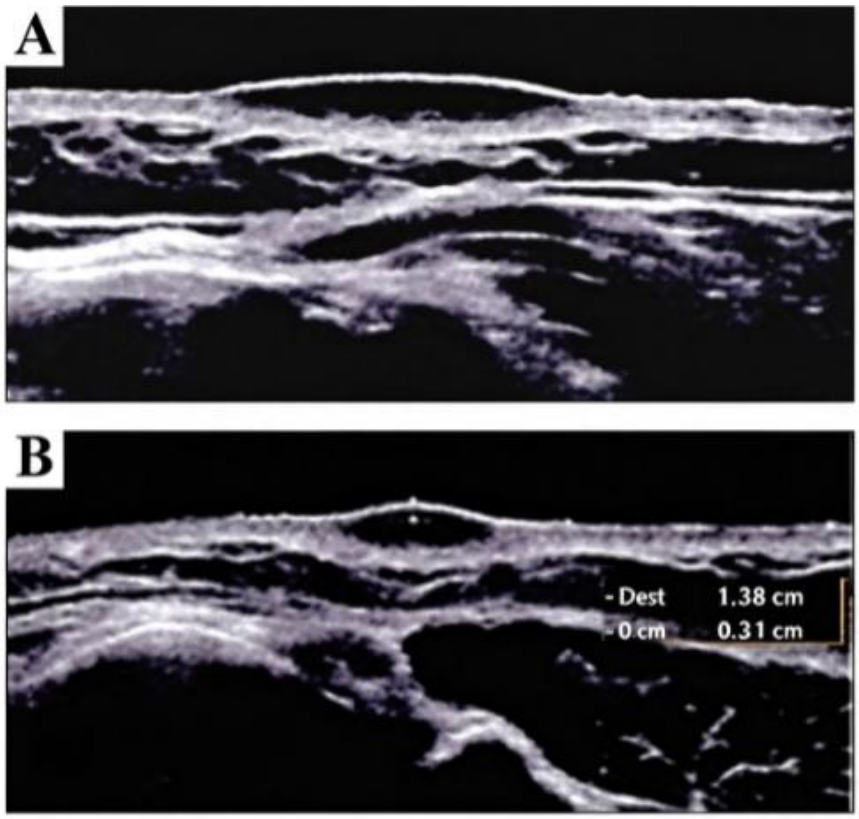


Fig. 1: Representative grayscale high-frequency ultrasound (HFUS) images of a keloid scar. (A) Long-axis section; (B) Short-axis section. "+" indicates the skin extension line; "x" indicates the measurement position for supralesional height. Images were acquired at 20 MHz. Representative image from one patient (29-year-old male, anterior chest lesion).

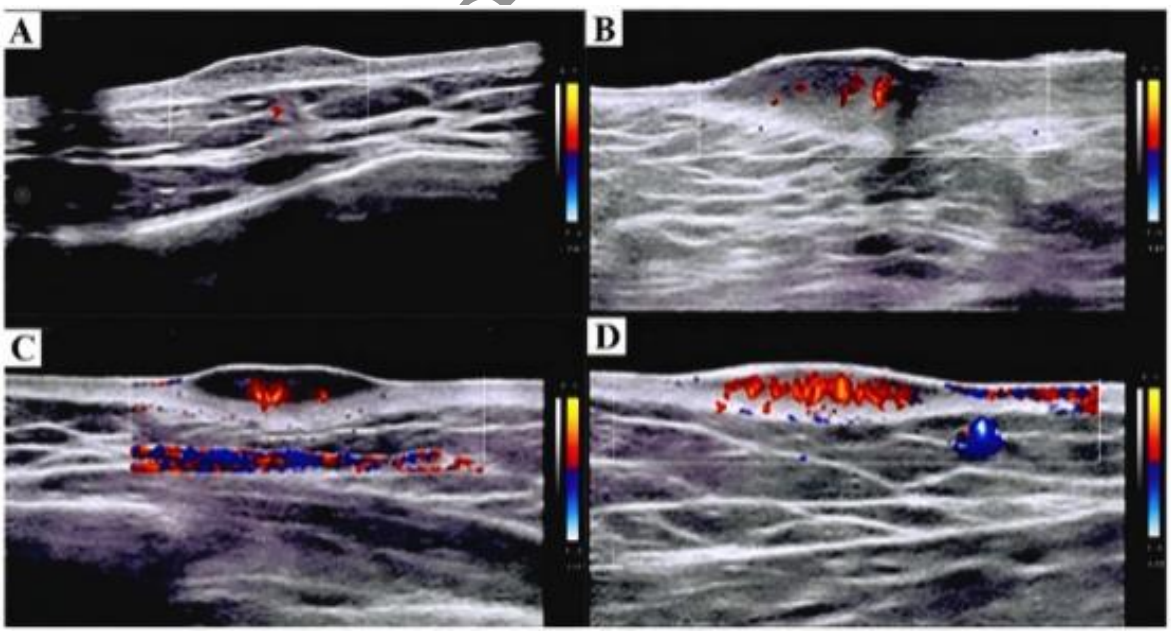


Fig. 2: Representative color micro-flow Doppler images demonstrating Adler vascularity grades. (A) Adler grade 0: no blood flow signal (27-year-old female, right knee lesion); (B) Adler grade I: 1–2 punctate or thin linear signals (54-year-old female, right lower abdominal lesion); (C) Adler grade II: 3–4 punctate signals or one obvious vessel (25-year-old female, anterior chest lesion); (D) Adler grade III: ≥5 punctate vessels or ≥2 longer vessels (29-year-old male, anterior chest lesion). Grades were defined according to Adler et al. (1990). Representative images from four patients.

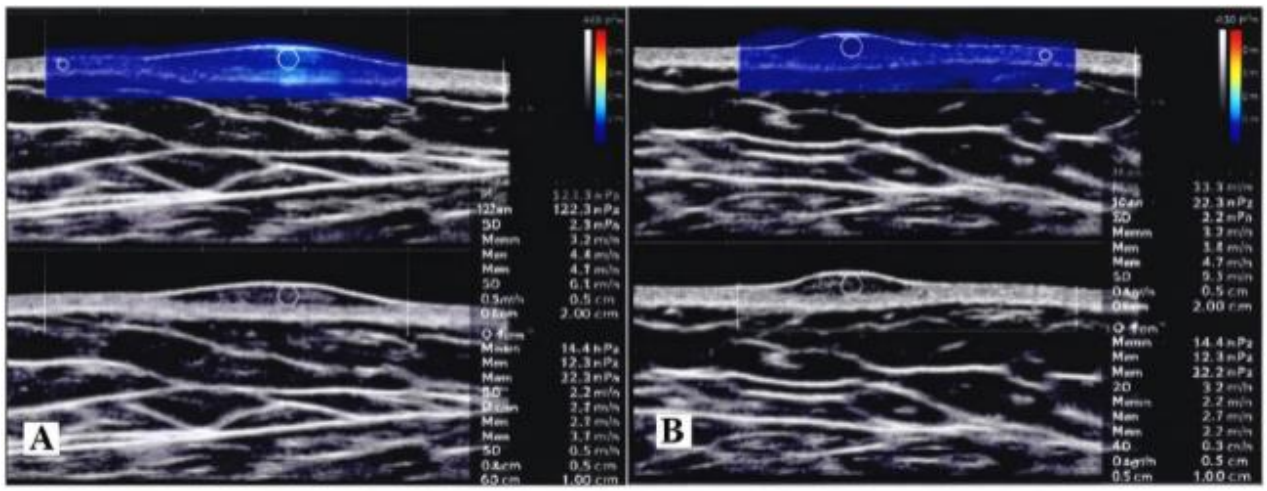


Fig. 3:Representative shear wave elastography (SWE) images of a keloid scar. (A) Long-axis section; (B) Short-axis section. The region of interest fully frames the lesion, and the Q-box diameter was adjusted according to scar thickness. Mean Young's modulus (kPa) was measured in the stiffest area. Representative image from one patient (29-year-old male, anterior chest lesion).

Shear wave elastography (SWE): The transducer was held perpendicular to the skin and lightly contacted the stand-off pad to avoid pre-compression artifacts. The color-coded region of interest was adjusted to encompass >50% of the scar area and include adjacent normal tissue. Once a stable elastogram was achieved, the image was frozen and the mean Young's modulus was measured in the stiffest area using a Q-box, with Q-box diameter tailored to scar thickness (representative SWE images are shown in figure 3. Averaged values from both sonographers were used for analysis to reduce random measurement error.

Determination of ultrasound-based activity typing cutoffs

The thickness cutoffs (5 mm and 7 mm) were exploratory thresholds based on clinical experience, pre-specified during protocol development rather than derived from post-hoc data exploration. Their selection was based on: (1) the general pharmacokinetic principle that drug concentration decreases with increasing diffusion distance through tissue; (2) previous ultrasound studies in keloids demonstrating an overall association between thickness and vascularity (Lobos *et al*., 2017) as well as treatment response (Cai *et al*., 2020); and (3) the practical utility of integer cutoffs for clinical memorization and application. The vascularity grade cutoff (Adler grade 0 vs. grades I–III) was based on prior classification standards (Adler *et al*., 1990). These cutoffs are exploratory, have not been directly validated by pharmacokinetic-pharmacodynamic studies, and require calibration and optimization in independent cohorts before implementation.

Treatment protocol

Compound betamethasone injection (Hangzhou MSD Pharmaceutical Co., Ltd., Guoyaozhunzi HJ20130187; commercially available as Diprospan in some regions) was drawn. The keloid lesion was disinfected according to standard procedures. The compound betamethasone preparation contained betamethasone dipropionate 5 mg/mL and betamethasone sodium phosphate 2 mg/mL. The drug was injected without routine dilution. The same senior dermatologist administered all injections and was not involved in any subsequent efficacy assessment or data measurement.

Injection volume calculation: Lesion surface area was calculated using the ellipse formula $S = \pi \times (a/2) \times (b/2)$, where a and b represent the longest and perpendicular diameters measured by high-frequency ultrasound. The injection volume was determined at 0.1–0.2 mL/cm² (surface area), with individualized adjustment based on lesion thickness: 0.1 mL/cm² for thickness <5 mm, 0.15 mL/cm² for thickness 5–7 mm and 0.2 mL/cm² for thickness >7 mm. The maximum total volume per session was 1.0 mL.

Standardization of injection technique: The same senior dermatologist administered all injections to eliminate inter-operator variability. A multi-point injection technique was used, distributing the total volume evenly across the central and peripheral zones of the lesion. Needle depth was stratified according to ultrasound-measured scar thickness: superficial injection (anterior one-third), mid-layer injection (middle one-third) and deep injection (posterior one-third), ensuring uniform drug distribution along the thickness gradient. Needle bevel orientation and injection pressure were kept consistent. Real-time ultrasound guidance was not used; however, injection sites and depths were pre-planned based on pre-procedural ultrasound measurements. The injection endpoint was defined as uniform blanching and mild tissue turgor of the lesion, avoiding over-injection that could lead to epidermal

blistering or focal atrophy. Treatment adherence was documented using a standardized electronic medical record system; all patients returned for treatment within the scheduled window (±3 days), with late returns or missed visits noted in the analysis. Treatment was repeated every 4 weeks for three sessions.

Follow-up and outcome assessment

Follow-up schedule and assessment time points

Week 4 was predefined as the early imaging response evaluation time point and the percentage reduction in high-frequency ultrasound thickness ($\Delta Th\%_4W$) was calculated using the formula: $\Delta Th\%_4W$ = (baseline thickness − Week 4 thickness) / baseline thickness × 100%.

Given the absence of an internationally recognized standard for defining early response at Week 4, the Week 12 overall thickness change rate ($\Delta Th\%12W$) ≥ 50% was used as the reference standard ("gold standard"). This threshold was selected pragmatically because it represents a clinically meaningful degree of thickness reduction and is consistent with the upper range of thickness improvement reported in previous ultrasound-based studies of intralesional corticosteroid therapy (Cai et al., 2020; Ardehali *et al*., 2007). However, it should be acknowledged that this 50% cutoff was not directly derived from those studies and requires further validation. Using ΔTh%□W as the test variable and Week 12 favorable efficacy as the state variable, a receiver operating characteristic (ROC) curve was constructed. The ROC analysis yielded an AUC of 0.867 (95% CI: 0.811–0.923). At the cutoff of ΔTh%□W = 30%, the Youden index reached its maximum (0.672), with a corresponding sensitivity of 81.5% (95% CI: 74.2%–87.3%) and specificity of 85.7% (95% CI: 78.9%–90.8%). Therefore, ΔTh%□W ≥ 30% was defined as favorable early response and <30% as poor early response.

It should be noted that both the cutoff determination and efficacy grouping were performed within the same dataset (n=150), which carries a risk of optimism bias. To partially correct for this bias, bootstrap internal validation (1,000 resampling iterations) was performed, yielding an optimism-corrected AUC of 0.851 (95% CI: 0.792 – 0.910), approximately 1.8% lower than the original AUC (0.867), suggesting an acceptable degree of overfitting. However, as this remains internal validation, the cutoff and the predictive model require external validation in an independent prospective cohort before clinical implementation.

Clinical rationale for binary classification

ΔTh%□W was analyzed both as a continuous and a categorical variable based on the following considerations: (1) clinical decision-making requires a clear threshold to determine whether a patient is responding to treatment, guiding decisions to continue the current regimen or promptly transition to combination therapy; (2) using the Week 12 overall thickness change rate as the reference standard, the optimal cutoff was determined by ROC analysis, balancing data-driven objectivity with maximum predictive concordance between the early imaging indicator and the longer-term efficacy endpoint; (3) the ultrasound-based activity stratification system is intended to facilitate rapid clinical risk stratification and a binary definition is more practical for bedside identification of high-risk patients who may benefit from early treatment modification.

Primary efficacy endpoint and recurrence definition

The Week 12 overall thickness change rate (ΔTh%□□W) served as the primary efficacy endpoint.

Recurrence was defined as any of the following after initial improvement at Week 12: increase in scar thickness (≥1 mm from Week 12), worsening of clinical symptoms (pruritus or pain), or increase in VSS total score ≥2 points from the best improvement achieved. Treatment durability and recurrence were observed through Week 24.

Subjective scale assessment

At each follow-up visit, scars were scored using the Vancouver Scar Scale (VSS) (Baryza and Baryza, 1995). The VSS score consists of four dimensions: pigmentation (0-2 points), vascular distribution (0-3 points), thickness (0-4 points) and softness (0-3 points). The total score ranges from 0 to 12 points, with higher scores indicating more severe scars. Changes in the VSS score were used to verify the external validity of the high-frequency ultrasound thickness measurement.

Rationale for the primary outcome measure

The Week 4 HFUS thickness change rate (ΔTh%□W) was selected as the primary outcome rather than Week 12 clinical improvement or Week 24 recurrence, based on three considerations. First, the need for a clinically actionable "time window." Keloid injection therapy is typically repeated at 4-week intervals (Murakami and Shigeki, 2024). In routine practice, physicians often assess efficacy only after 2–3 injections (i.e., Week 8–12), meaning poor responders have already received at least two "ineffective" injections before any treatment modification occurs. By identifying potential non-responders at Week 4—the interval between the first and second injections—clinicians could adjust treatment at the next session, advancing the decision point by 4–8 weeks. Second, the association between early ultrasound response and longer-term efficacy. Although Week 24 recurrence is the ultimate endpoint for assessing treatment durability, it occurs months after treatment cessation and offers little guidance for modifying the ongoing therapeutic regimen. This study tests whether Week 4 ultrasound response predicts Week 12 overall efficacy (the conventional time point for the first comprehensive clinical assessment), establishing a

In Press

predictive chain from "early indicator" to "mid-term endpoint." The Week 4 favorable response group achieved an 83.7% (77/92) Week 12 overall favorable rate, providing preliminary empirical support for this rationale. Third, ultrasound thickness as a surrogate for clinical endpoints. Previous studies have demonstrated correlations between ultrasound-derived thickness changes and clinical scar assessments (Xu *et al*., 2024; Cai *et al*., 2020). The correlation coefficient between changes in the VSS thickness subscale and ultrasound-derived thickness reduction was r = 0.814, confirming that ultrasound thickness reduction effectively reflects scar softening and symptomatic improvement. The ultimate goal of this study is not to replace clinical judgment with ultrasound, but to provide an objective tool that can detect treatment response earlier and more sensitively than visual inspection or subjective scoring, enabling timely therapeutic adjustments at critical decision nodes.

Laboratory indicators

Before treatment and after 3 months of treatment, 2 mL of peripheral venous blood was collected, allowed to clot at room temperature for 30 min and centrifuged at 3000 rpm for 15 min. Serum was separated, aliquoted and stored at −80°C until analysis. Serum levels of TGF-β□ (Cat# KE00021), EGF (Cat# KE00045), 5-HT (Cat# KE00123) and bFGF (Cat# KE00067) were quantified using commercial enzyme-linked immunosorbent assay (ELISA) kits (Shanghai Keaibo Biotechnology Co., Ltd., China) according to the manufacturer's instructions. All samples were run in duplicate and the average value was used for analysis. The detection ranges were: TGF-β_1 15.6–1000 pg/mL, EGF 3.9–250 pg/mL, 5-HT 0.78–50 ng/mL and bFGF 7.8–500 pg/mL. Intra-assay coefficients of variation (CVs) were <8%, and inter-assay CVs were <12%, as reported by the manufacturer and verified in the present laboratory. All samples were measured in the same batch to eliminate inter-batch variability.

Molecular analysis (exploratory subgroup)

For exploratory molecular analyses, a subgroup of 40 patients (20 from the favorable early-response group and 20 from the poor early-response group) was included. These patients were not randomly selected; rather, they were a convenience subgroup who underwent clinically indicated surgical excision of keloid lesions (e.g., for symptomatic relief or cosmetic improvement) and provided consent for molecular analysis of the excised tissue. No additional biopsy was performed solely for research purposes. Importantly, all tissue samples used for molecular analysis were obtained from pre-treatment, non-injected lesions. Specifically, tissue was collected either before the first intralesional corticosteroid injection or from lesions that had not received any prior intralesional injection. Lesions that had already received intralesional corticosteroid therapy before tissue collection were excluded from the molecular analysis.

This molecular analysis should therefore be considered exploratory and hypothesis-generating. Because the subgroup was derived from surgical excision specimens rather than random biopsies, it is subject to selection bias and may not be representative of the entire keloid cohort. The findings require further validation in larger prospective studies with random or consecutive sampling.

The epidermis and subcutaneous fat were excised to isolate the dermal layer. Total RNA was extracted from keloid tissues using TRIzol® Reagent (Invitrogen, USA). RNA concentration and purity were assessed using a NanoDrop 2000 spectrophotometer (Thermo Scientific, USA); samples were included only if they had an A260/A280 ratio between 1.8 and 2.1. A total of 1 μg of RNA was reverse-transcribed using M-MLV Reverse Transcriptase (Promega, USA) with random primers in a 20 μL reaction volume under the following conditions: 25°C for 5 min, 42°C for 60 min and 70°C for 5 min for enzyme inactivation.

Real-time quantitative PCR was performed using Luna Universal qPCR Master Mix (New England Biolabs, USA) on a LightCycler® 480 II system (Roche, Switzerland). The reaction mixture (10 μL total volume) contained 2 μL cDNA template, 0.5 μL each of the forward and reverse primers (final concentration 250 nM), 5 μL Master Mix, and 2 μL nuclease-free water. All reactions were performed in triplicate, with no-template controls included as negative controls. GAPDH was used as the internal reference gene for mRNA (RUNX2), and U6 was used for miRNA (miR-205-5p). The thermal cycling protocol consisted of: initial denaturation at 95°C for 2 min, followed by 40 cycles of denaturation at 95°C for 15 s and combined annealing/extension at 60°C for 30 s. After cycling, melting curve analysis was performed (65°C to 95°C at 0.5°C/s) to verify amplification specificity. Amplification efficiency was determined using a standard curve generated from 10-fold serial dilutions of cDNA (6 dilutions, 10^{o} to 10^{-5}). All primer pairs exhibited amplification efficiencies between 90% and 110% (RUNX2: 96.8%, GAPDH: 98.2%; miR-205-5p: 94.5%, U6: 97.1%), with correlation coefficients R^2 >0.99 for all standard curves, satisfying the applicability criteria for the $2^{-\Delta\Delta CT}$ method. Relative expression levels were calculated using the $2^{-}\Delta\Delta CT$ method, with the control group (Adler grade 0 group) as the calibrator. Primer sequences are shown in table 1.

Selection bias in the molecular subgroup: The 40 patients included in the molecular substudy had tissue samples obtained from clinically indicated keloid excision surgery rather than random biopsies.

Article in Press

Table 1: Primer sequences.

Gene name	Primer direction	Primer sequence (5'→3')
RUNX2	Forward	CGGAATGCCTCTGCTGTTAT
	Reverse	TTCCCGAGGTCCATCTACTG
U6	Forward	CTCGCTTCGGCAGCACA
	Reverse	AACGCTTCACGAATTTGCGT
GAPDH	Forward	TGTTCGTCATGGGTGTGAAC
	Reverse	ATGGCATGGACTGTGGTCAT
miR-205-5p	Forward	TCCTTCATTCCACCGGAGTC
	Reverse	AACATGTACAGTCCATGGATG

Table 2: Intra-observer and inter-observer reliability of high-frequency ultrasound measurements.

Parameter	Intra-observer ICC (95% CI)*	Inter-observer ICC (95% CI)†
Scar thickness (mm)	0.948 (0.922–0.974)	0.936 (0.908–0.963)
Scar short-axis length (mm)	0.932 (0.894–0.958)	0.915 (0.876–0.946)
Supralesional height (mm)	0.910 (0.862–0.942)	0.898 (0.852–0.932)
Short-axis Young's modulus mean (kPa)	0.924 (0.884–0.952)	0.912 (0.868–0.944)
Adler Doppler vascularity grade	0.936 (0.900–0.966)‡	0.921 (0.872–0.948)‡

Note: * Intra-observer ICC: two measurements by the same sonographer at ≥24-hour intervals, n=20; † Inter-observer ICC: independent measurements by two sonographers on the same set of lesions, n=20; ‡ Weighted Kappa coefficient was used for Adler vascularity grade.

Compared with the 110 patients who did not undergo surgical excision, the surgical subgroup exhibited the following characteristics: (1) larger lesions (mean diameter 8.7±2.4 cm vs. 6.5±2.1 cm in the non-surgical group, P=0.008); (2) more severe symptoms (baseline VSS score 9.8±1.5 vs. 8.6±1.8, P=0.012); (3) longer disease duration (median 4.2 years vs. 2.8 years, P=0.021); and (4) a higher proportion of the high-activity type (35.0% in the surgical subgroup vs. 6.4% in the non-surgical group, P<0.001). These differences indicate that the surgical subgroup represents a keloid subpopulation with greater disease burden and higher clinical activity.

This selection bias has two opposing effects on the molecular findings. On the favorable side, because the surgical subgroup includes a higher proportion of high-activity and high-VSS lesions, the differences in miR-205-5p and RUNX2 expression are more readily detected as statistically significant, enhancing the statistical power of the molecular correlation analysis. On the downside, the surgical subgroup cannot represent "mild" or "quiescent" lesions across the full keloid spectrum, and it remains unclear whether the observed molecular expression pattern (low miR-205-5p/high RUNX2) also exists in non-surgical patients. Moreover, the overrepresentation of the high-activity type in the surgical subgroup (35.0% vs. 6.4%) suggests that the molecular correlations observed here may overestimate the strength of the association between ultrasound parameters and molecular markers in the general keloid population. Therefore, these molecular findings should be considered exploratory associations within a high-activity keloid subpopulation and their generalizability to the broader keloid population requires validation in random biopsy-based or prospective cohort studies.

Quality control

The following blinding strategy was implemented: (1) two independent sonographers were blinded to group assignment and clinical outcomes and were not involved in treatment decisions; (2) two dermatologists performing VSS evaluations were blinded to group allocation and treatment procedures; (3) the injecting physician could not be blinded to treatment allocation due to the need for real-time dose and depth adjustment, but was not involved in any efficacy assessment or data measurement.

To ensure data integrity: (1) All ultrasound examinations were performed independently by two sonographers blinded to efficacy grouping; (2) Each sonographer measured all parameters in triplicate and averaged the values, with inter-observer ICC calculated to assess agreement; (3) All intralesional injections were administered by the same designated physician using standardized techniques; (4) Data were independently entered and cross-verified by two researchers, with specific reasons documented for any missing values; (5) Patients received advance telephone reminders and transportation subsidies to minimize attrition; (6) A maximum of two missed visits was permitted; patients exceeding this threshold were excluded from the final analysis.

Statistical analysis

SPSS 27.0 was used. Independent-samples t-tests, χ^2 tests and linear mixed-effects models were used for between-group comparisons and repeated thickness measurements.

Multivariable logistic regression with forward stepwise selection identified independent predictors of poor early response; discrimination was assessed by ROC analysis and bootstrap internal validation. Reliability was evaluated using intraclass correlation coefficients and weighted Kappa, and two-sided $P < 0.05$ was considered significant.

RESULTS

Ultrasound measurement reliability

Reliability testing before formal enrollment and during the study showed intra-observer and inter-observer ICCs> 0.90 for all ultrasound parameters (Table 2), indicating excellent repeatability and agreement. Scar thickness showed the highest reliability, with an intra-observer ICC of 0.948 (95% CI: 0.922–0.974) and inter-observer ICC of 0.936 (95% CI: 0.908–0.963). The weighted Kappa coefficients for Adler vascularity grade were 0.936 (intra-observer) and 0.921 (inter-observer), also indicating excellent agreement.

Baseline characteristics of the two groups

A total of 150 keloid patients were enrolled. According to the Week-4 early imaging response ($\Delta Th\%_4W$), 92 patients (61.3%) were classified as the favorable early-response group ($\Delta Th\%_4W \geqslant 30\%$) and 58 patients (38.7%) as the poor early-response group ($\Delta Th\%_4W < 30\%$).

The two groups did not differ significantly in sex, body mass index, smoking history, drinking history, or scar duration (all $P > 0.05$). However, the poor early-response group was significantly older than the favorable early-response group (mean difference = 7.4 years, 95% CI: 3.8–11.0, $P = 0.018$), had a higher proportion of dry skin (risk difference = 18.3%, 95% CI: 3.8%–32.8%, $P = 0.041$), higher baseline VSS total scores (mean difference = 1.2 points, 95% CI: 0.3–2.1, $P = 0.027$), a higher proportion of lesions with maximum diameter ≥4 cm (risk difference = 16.9%, 95% CI: 1.2%–32.6%, $P = 0.045$) and a higher prevalence of concomitant skin diseases (risk difference = 14.6%, 95% CI: 1.8%–27.4%, $P = 0.032$) (Table 3, Fig. 4).

Ultrasound and blood flow signal characteristics

No significant differences were observed between the two groups in echo distribution or echo type (both $P > 0.05$). The favorable early-response group had a significantly lower proportion of positive intralesional vascularity compared with the poor early-response group (risk difference = -34.7%, 95% CI: -49.8% to -19.6%, $P < 0.001$), as well as smaller scar thickness (mean difference = -4.3 mm, 95% CI: -5.6 to -3.0, $P < 0.001$), shorter scar short-axis length (mean difference = -3.8 mm, 95% CI: -5.2 to -2.4, $P < 0.001$), lower supralesional height (mean difference = -1.2 mm, 95% CI: -1.9 to -0.5, $P = 0.038$) and lower mean short-axis Young's modulus (mean difference = -8.6 kPa, 95% CI: -15.2 to -2.0, $P = 0.033$) (Table 4, Fig. 5).

Treatment response and thickness reduction

At Week 4, the overall $\Delta Th\%\square W$ was 28.5% ± 8.9%. The favorable early-response group achieved a significantly greater thickness reduction than the poor early-response group (mean difference = 18.6%, 95% CI: 15.2%–22.0%, $P < 0.001$). As baseline scar thickness increased, $\Delta Th\%_4W$ declined progressively. Within the same thickness stratum, $\Delta Th\%_4W$ in the blood flow-positive group was lower than that in the blood flow-negative group (<5 mm: 31.5% vs. 38.2%; 5–7 mm: 22.4% vs. 29.8%; >7 mm: 15.6% vs. 24.1%) (Table 5, Fig. 6).

At Week 12, the overall thickness change rate ($\Delta Th\%\square\square W$) was 45.3% ± 12.6%, with 98 patients (65.3%) achieving favorable overall efficacy ($\Delta Th\%\square\square W \geq 50\%$). The linear mixed-effects model revealed a significant group × time interaction ($F = 8.342$, $P < 0.001$). The adjusted mean difference in thickness between the two groups across all time points was -3.4 mm (95% CI: -4.6 to -2.2, $P < 0.001$). The adjusted mean differences at Week 4, Week 8 and Week 12 were -4.0 mm (95% CI: -5.2 to -2.8), -3.4 mm (95% CI: -4.6 to -2.2) and -3.4 mm (95% CI: -4.6 to -2.2), respectively (Table 6).

Concordance between subjective VSS and objective ultrasound parameters

At baseline, the two groups did not differ significantly in total VSS score (8.9 ± 1.7 vs. 9.1 ± 1.8, $P = 0.494$). At Week 4, the total VSS score in the favorable early-response group decreased to 5.6 ± 1.3, which was significantly lower than that in the poor early-response group (7.4 ± 1.5, $P < 0.001$).

The change in the VSS thickness subscale showed a strong positive correlation with the ultrasound-derived thickness change rate ($r = 0.81$, 95% CI: 0.74–0.89, $P < 0.001$) (Table 7).

Laboratory indicators

Before treatment, serum levels of EGF, TGF-β_1, 5-HT, and bFGF did not differ significantly between the two groups (all $P > 0.05$). After treatment, all four indicators in the favorable early-response group were significantly lower than those in the poor early-response group: EGF (155.94 ± 10.93 ng/L vs. 175.50 ± 11.28 ng/L, mean difference = -19.56 ng/L, 95% CI: -25.12 to -14.00, $P < 0.001$); TGF-β_1 (201.56 ± 12.48 ng/L vs. 232.72 ± 13.94 ng/L, mean difference = -31.16 ng/L, 95% CI: -39.62 to -22.70, $P < 0.001$); 5-HT (107.45 ± 12.19 ng/L vs. 125.37 ± 12.46 ng/L, mean difference = -17.92 ng/L, 95% CI: -24.28 to -11.56, $P < 0.001$); bFGF (78.26 ± 8.73 ng/L vs. 96.15 ± 9.42 ng/L, mean difference = -17.89 ng/L, 95% CI: -23.15 to -12.63, $P < 0.001$) (Table 8, Fig. 7).

nicle.in Press

Table 3: Comparison of baseline characteristics between the two groups.

Characteristic	Favorable early-response group (n=92)	Poor early-response group (n=58)	t/χ^2	Effect size (95% CI)	*P*-value
Age (years)	34.20 ± 8.50	41.60 ± 10.20	t=4.82	MD=7.40 (3.80–11.00)	0.018
Male [n (%)]	48 (52.2)	30 (51.7)	χ^2=0.003	RD=0.5% (-16.8%–17.7%)	0.956
BMI (kg/m²)	22.80 ± 2.50	23.10 ± 2.70	t=0.71	MD=0.30 (-0.53–1.13)	0.482
Disease duration (months)	28.40 ± 15.20	31.60 ± 18.40	t=1.14	MD=3.20 (-2.37–8.77)	0.258
Lesion longest diameter (cm)	6.40 ± 2.00	7.20 ± 2.30	t=1.96	MD=0.80 (-0.01–1.61)	0.052
Lesion ≥4 cm [n (%)]	32 (34.8)	30 (51.7)	χ^2=4.02	RD=16.9% (1.2%–32.6%)	0.045
Dry skin [n (%)]	18 (19.6)	22 (37.9)	χ^2=4.19	RD=18.3% (3.8%–32.8%)	0.041
Concomitant skin diseases [n (%)]	12 (13.0)	16 (27.6)	χ^2=4.61	RD=14.6% (1.8%–27.3%)	0.032
Baseline VSS total score (points)	8.90 ± 1.70	9.10 ± 1.80	t=2.23	MD=1.20 (0.27–2.13)	0.027

Note: MD, mean difference; RD, risk difference.

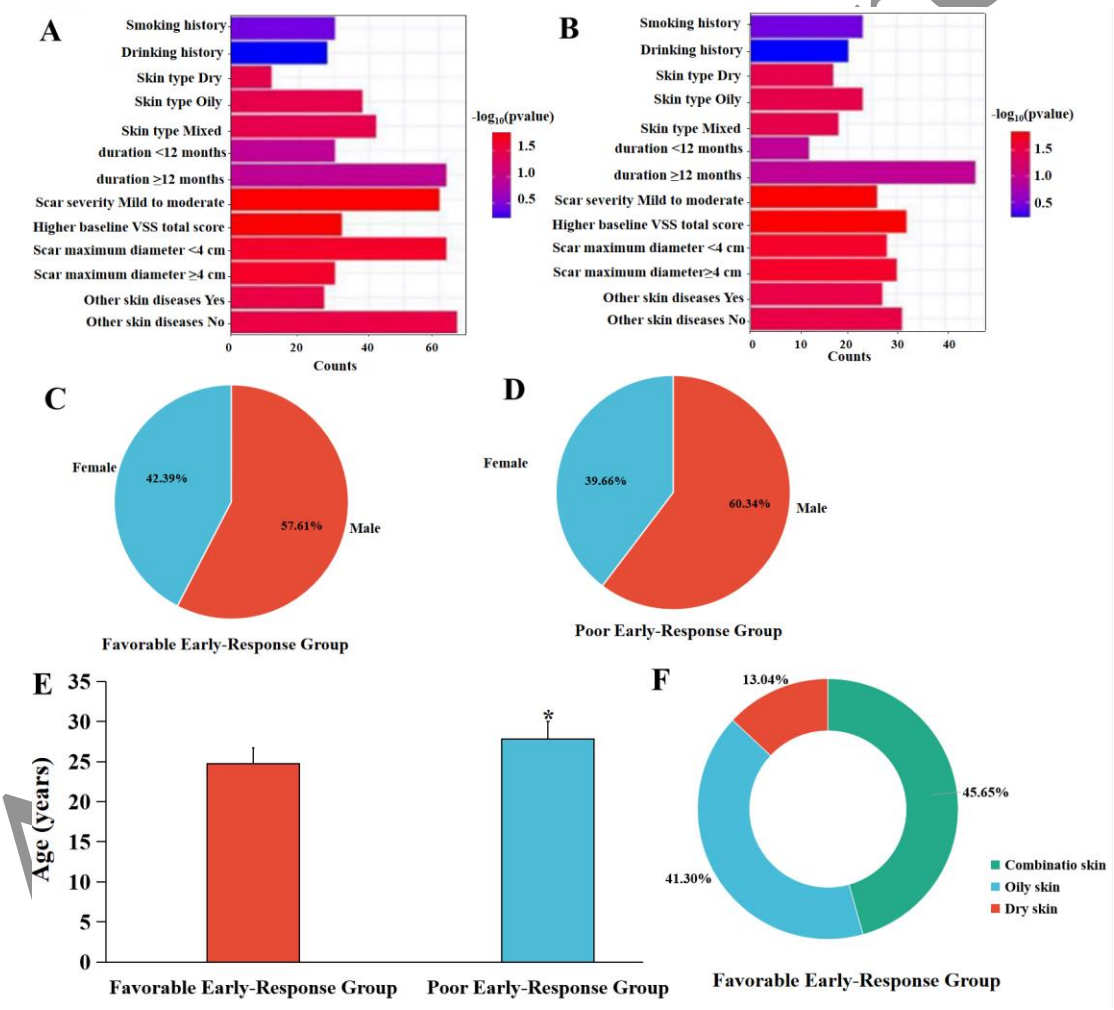


Fig. 4: Comparison of baseline characteristics between the favorable and poor early-response groups. (A) Clinical features of the favorable early-response group; (B) clinical features of the poor early-response group; (C) sex distribution in the favorable early-response group; (D) sex distribution in the poor early-response group; (E) age comparison between the two groups; (F) skin type in the favorable early-response group. Data are shown as mean ± SD. *P < 0.05 vs. the favorable early-response group (independent-samples t-test for continuous variables; X^2 test for categorical variables). n = 92 (favorable) and n = 58 (poor).

Table 4: Comparison of ultrasound and blood flow signal characteristics between the two groups [n (%)].

Ultrasound characteristic		Favorable early-response group (n=92)	Poor early-response group (n=58)	χ^2	*P*
Echo distribution [n (%)]	Heterogeneous	24 (26.1)	23 (39.7)	3.04	0.081
	68 (73.9)	35 (60.3)	68 (73.9)		
Echo/signal intensity [n (%)]	Hypoechoic	80 (87.0)	44 (75.9)	3.06	0.080
	Isoechoic	12 (13.0)	14 (24.1)		
Vascularity [n (%)]	Absent blood flow	51 (55.4)	12 (20.7)	7.63	<0.001†
	Positive blood flow	41 (44.6)	46 (79.3)		

Note: Echo distribution, echo/signal intensity and vascularity were analyzed as separate ultrasound characteristics. A total of 3 independent tests were performed in this table. The significance threshold after Bonferroni correction was $\alpha' = 0.017$. † indicates $P < 0.017$ after correction.

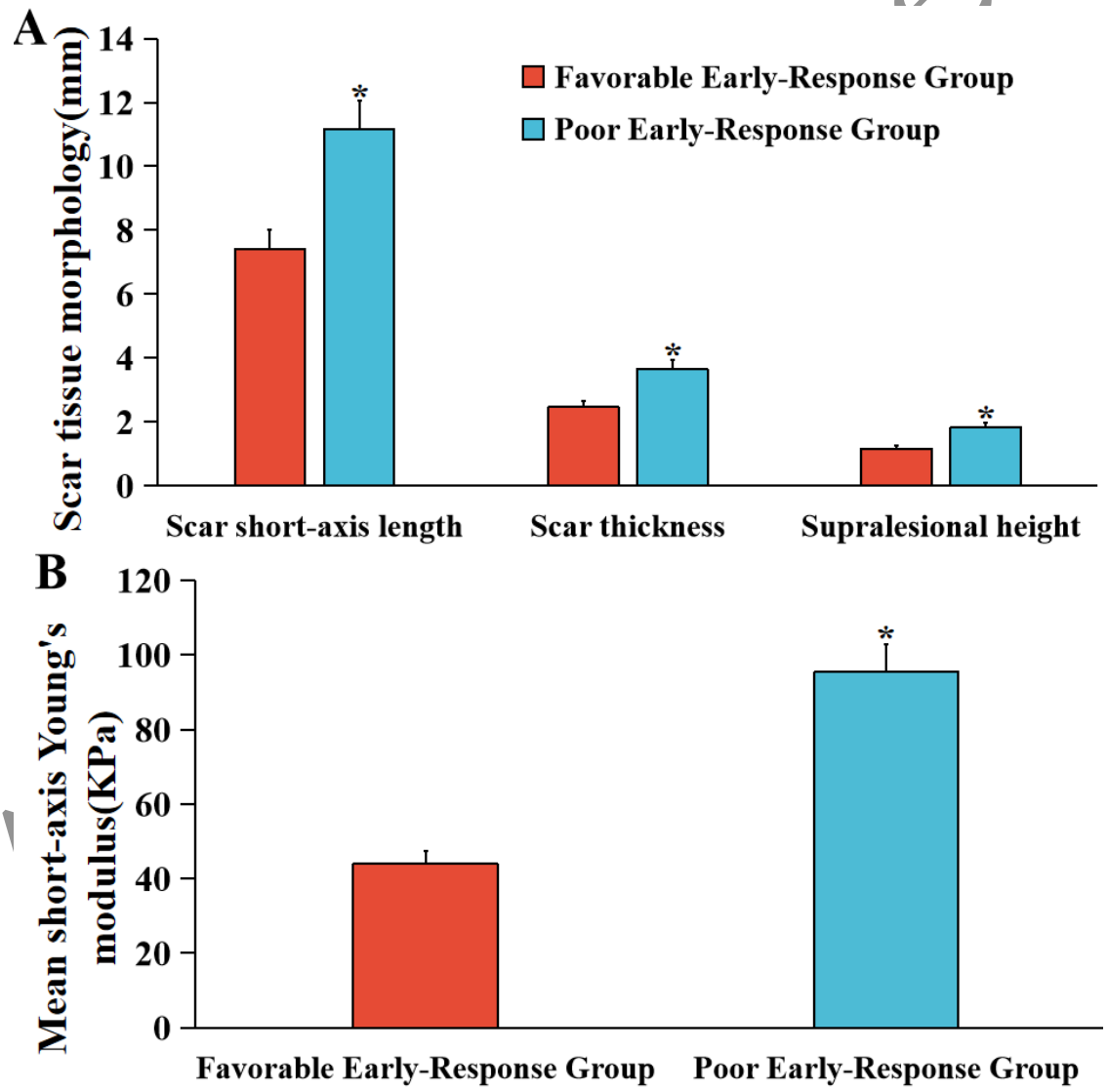


Fig. 5: Comparison of ultrasound measurement parameters between the two groups. (A) Scar tissue morphology; (B) mean short-axis Young's modulus. Data are shown as mean ± SD. *P < 0.05 vs. the favorable early-response group (independent-samples t-test). n = 92 (favorable) and n = 58 (poor).

Table 5: Comparison of ultrasound and blood flow signal characteristics between the two groups.

Baseline blood flow grade	No. of cases (n)	Percentage thickness reduction at 4 weeks ($\Delta Th\%_4$ W, %)	Favorable early-response group (n, %)	Poor early-response group (n, %)
Adler Grade 0 (No blood flow)	68	34.5 ± 6.2	48 (70.6%)	20 (29.4%)
Adler Grade I	42	27.1 ± 5.8	24 (57.1%)	18 (42.9%)
Adler Grade II	28	20.3 ± 4.9	14 (50.0%)	14 (50.0%)
Adler Grade III	12	14.2 ± 3.5	6 (50.0%)	6 (50.0%)
Total	150	28.5 ± 8.9	92 (61.3%)	58 (38.7%)

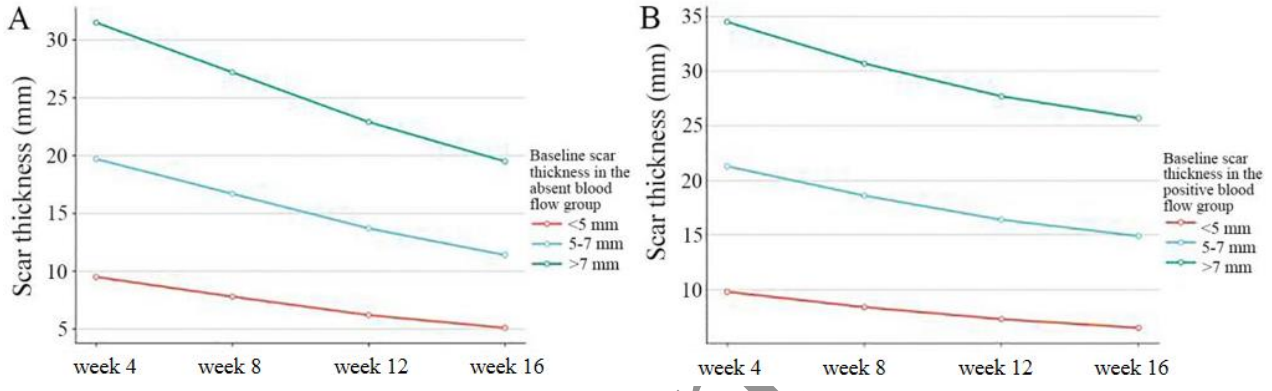


Fig. 6: Trends of Δ Th%,W stratified by baseline thickness and blood flow status. (A) Scar thickness changes across baseline thickness strata in the blood flow-negative group (Adler grade 0) at Weeks 4, 8, 12 and 16; (B) scar thickness changes across baseline thickness strata in the blood flow-positive group (Adler grades I‑III) at Weeks 4, 8, 12 and 16. Data are estimated marginal means from linear mixed-effects models; error bars indicate 95% CIs. n = 92 (favorable) and n = 58 (poor).

Table 6: Adjusted marginal means of scar thickness over time in the favorable and poor early-response groups.

Time point	Favorable early-response group (n=92)	Poor early-response group (n=58)	Adjusted mean difference (95% CI)	*P*-value
Baseline	8.50 ± 0.40	12.80 ± 0.50	-4.30 (-5.60 to -3.00)	<0.001
Week 4	5.60 ± 0.30	9.60 ± 0.50	-4.00 (-5.20 to -2.80)	<0.001
Week 8	4.80 ± 0.30	8.20 ± 0.50	-3.40 (-4.60 to -2.20)	<0.001
Week 12	4.10 ± 0.30	7.50 ± 0.50	-3.40 (-4.60 to -2.20)	<0.001

Note: Data are adjusted marginal means (95% CI) estimated from linear mixed-effects models with group, time and group × time interaction as fixed effects and individual patients as random intercepts. The group × time interaction was significant (F = 8.342, P < 0.001). Adjusted mean differences between groups at each time point are shown. No stratification by blood flow grade was performed in this table.

Table 7: Comparison of VSS scores between the two groups (±s, points).

Indicator		Favorable early-response group (n=92)	Poor early-response group (n=58)	*t*	*P*
VSS total score	Pre-treatment	8.9±1.7	9.1±1.8	0.69	0.494
	Post-treatment	5.6±1.3	7.4±1.5	7.78	<0.001†
Thickness subscale VSS-	Pre-treatment	2.8±0.6	2.9±0.7	0.93	0.353
	Post-treatment	1.5±0.5	2.2±0.6	7.72	<0.001†

Note: A total of 4 independent tests were performed in this table. The significance threshold after Bonferroni correction was $\alpha' = 0.0125$. † indicates P < 0.0125 after correction.

Table 8: Comparison of laboratory indicators between the two groups ($\bar{x}\pm s$, ng/L).

Serum indicators		Favorable early-response group (n=92)	Poor early-response group (n=58)	Mean difference (95% CI)	*P*
EGF	Pre-treatment	230.78±15.37	232.05±15.33	—	0.623
	Post-treatment	155.94±10.93	175.50±11.28	-19.56 (-25.12 to -14.00)	<0.001†
TGF-β_1	Pre-treatment	422.49±20.28	422.13±21.32	—	0.918
	Post-treatment	201.56±12.48	232.72±13.94	-31.16 (-39.62 to -22.70)	<0.001†
5-HT	Pre-treatment	220.63±25.27	219.58±24.70	—	0.803
	Post-treatment	107.45±12.19	125.37±12.46	-17.92 (-24.28 to -11.56)	<0.001†
bFGF (ng/L)	Pre-treatment	157.38 ± 12.51	158.22 ± 13.04	—	
	Post-treatment	78.26 ± 8.73	96.15 ± 9.42	-17.89 (-23.15 to -12.63)	

Note: A total of 8 independent tests were performed in this table. The significance threshold after Bonferroni correction was $\alpha' = 0.006$. † indicates $P < 0.006$ after correction.

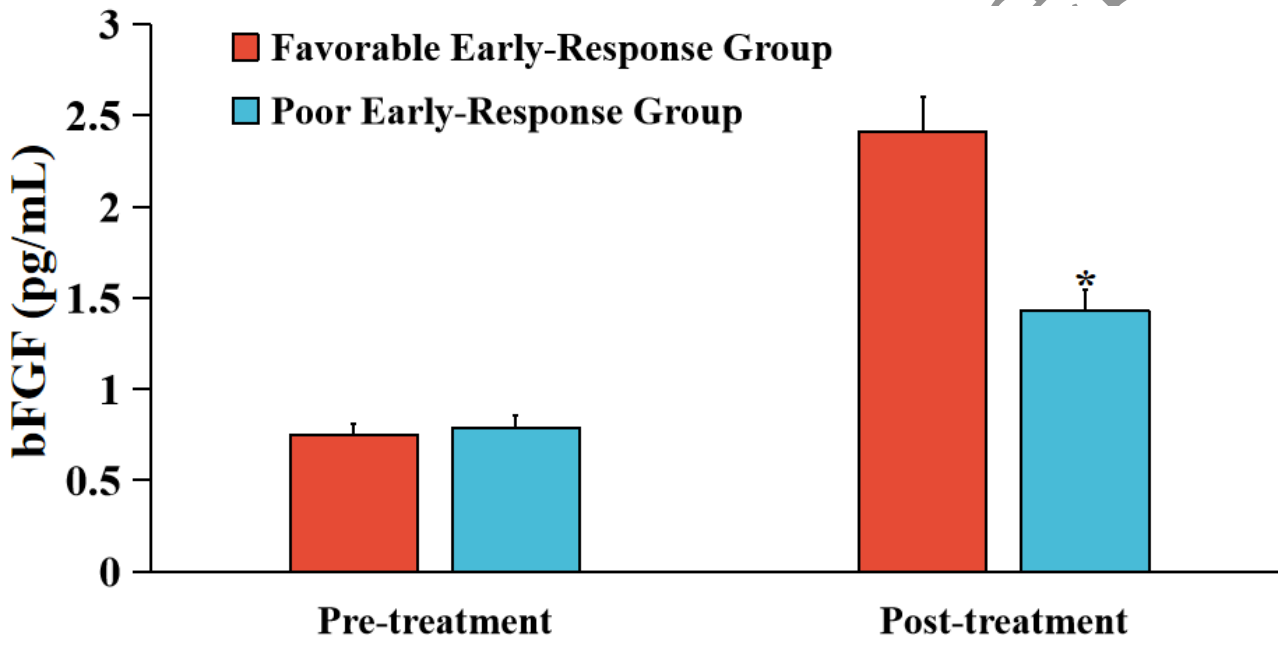


Fig. 7: Comparison of bFGF levels before and after treatment between the two groups. Data are shown as mean ± SD. *P < 0.05 vs. the favorable early-response group (independent-samples t-test). n = 92 (favorable) and n = 58 (poor).

Table 9: Comparison of baseline characteristics between the subgroup patients and the total population.

Characteristics	Total population (n=150)	Molecular marker subgroup (n=40)	χ^2/ t	*P*
Proportion of poor early-response group	58 (38.7%)	20 (50.0%)	1.84	0.175
Baseline scar thickness (mm)	10.2 ±3.5	10.5 ±3.8	-0.52	0.604
Proportion of positive blood flow (Adler I-Ⅲ)	82 (54.7%)	18 (45.0%)	1.33	0.249

Correlation analysis of baseline ultrasound parameters and molecular markers (exploratory subgroup analysis)

An exploratory molecular subgroup of 40 patients (20 from the favorable early-response group and 20 from the poor early-response group) was included. These patients were not randomly selected; they were a convenience subgroup who underwent clinically indicated surgical excision and consented to molecular analysis of the excised tissue. All tissue samples were obtained from pre-treatment, non-injected lesions. Compared with the total cohort, this subgroup showed no significant differences in several key baseline characteristics (all P > 0.05; Table 9); however, as detailed in Section Molecular Analysis (Exploratory Subgroup), the surgical subgroup exhibited a higher disease burden and a greater proportion of high-activity lesions, indicating a non-negligible selection bias. Therefore, the molecular findings should be regarded as exploratory associations within a surgical subgroup rather than as representative findings for the entire keloid population.

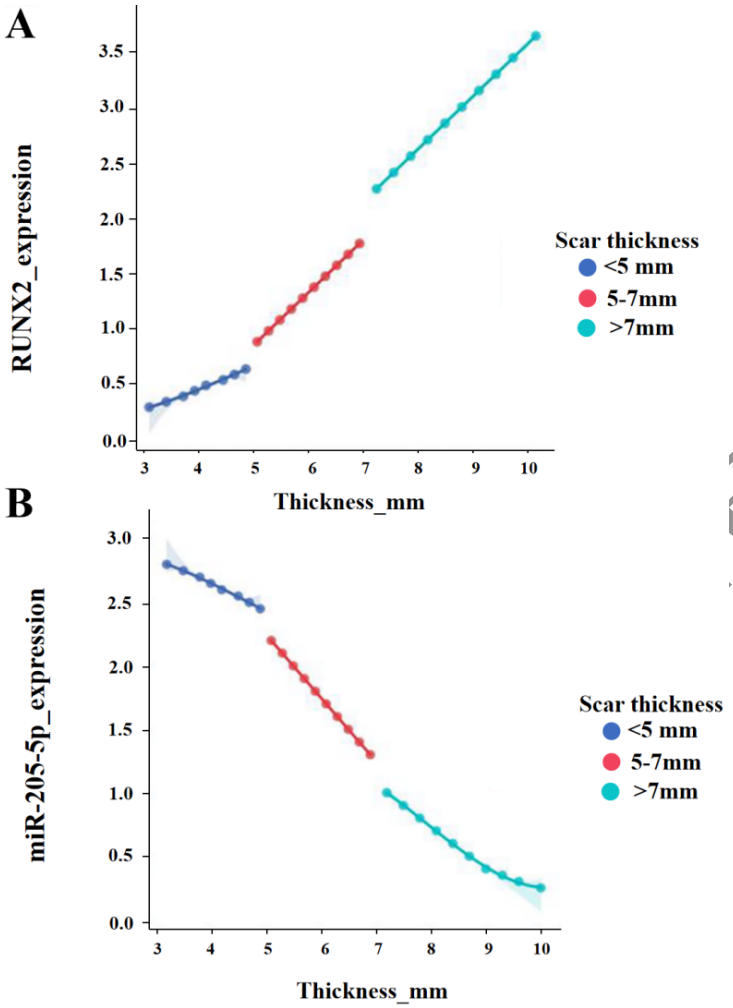


Fig. 8. Correlation scatter plots between baseline scar thickness and molecular marker expression levels. (A) Baseline scar thickness showed a significant positive correlation with RUNX2 expression (Spearman r = 0.921, P < 0.001) Linear regression equation: Y = −0.68 + 0.44X; (B) baseline scar thickness showed a significant negative correlation with miR-205-5p expression (Spearman r = −0.873, P < 0.001). Linear regression equations are shown. n = 40.Linear regression equation: Y = 3.41 − 0.32X.

Table 10: Spearman correlation matrix between molecular markers and ultrasound parameters.

Variable	miR-205-5p expression	RUNX2 expression	Baseline scar thickness	Adler vascularity grade
miR-205-5p expression	1.000	—	—	—
RUNX2 expression	-0.682 (-0.798 to -0.512)	1.000	—	—
Baseline scar thickness	-0.873 (-0.942 to -0.736)	0.921 (0.839–0.962)	1.000	—
Adler vascularity grade	-0.625 (-0.784 to -0.412)	0.643 (0.436–0.798)		

Association with scar thickness: Correlation analysis showed a significant negative correlation between baseline scar thickness and miR-205-5p expression (r = -0.87, 95% CI: -0.94 to -0.74, P < 0.001, n = 40) and a significant positive correlation with RUNX2 expression (r = 0.92, 95% CI: 0.84–0.96, P < 0.001, n = 40) (Fig. 8). miR-205-5p and RUNX2 expression also showed a significant negative correlation (r = -0.68, 95% CI: -0.80 to -0.51, P < 0.001) (Table 10).

Association with blood flow grade: miR-205-5p expression was significantly higher in the blood flow-negative group (Adler grade 0) than in the blood flow-positive group (Adler grades I–III) (mean difference = 0.63, 95% CI: 0.44–0.82, P < 0.001), while RUNX2 expression was significantly lower (mean difference = -0.69, 95% CI: -0.94 to -0.44, P < 0.001) (Table 11).

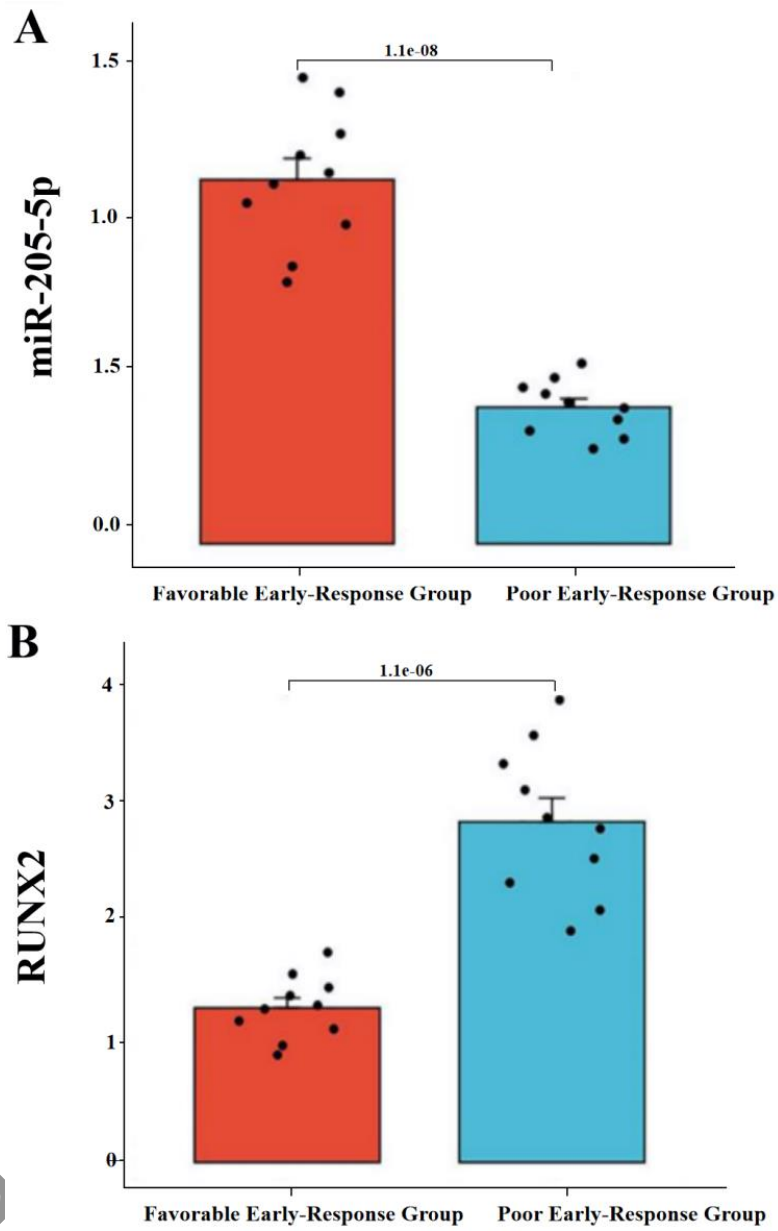


Fig. 9: Comparison of miR-205-5p and RUNX2 expression levels between the two groups. (A) miR-205-5p expression; (B) RUNX2 expression. Data are shown as mean ± SD. *P < 0.05 vs. the favorable early-response group (independent-samples t-test). n = 20 per group.

Table 11: Comparison of molecular marker expression levels across different blood flow signal grading groups.

Blood flow grade	Sample size (n)	miR-205-5p expression (Mean ±SD)	RUNX2 expression (Mean ±SD)
Adler Grade 0 (Absent blood flow)	22	1.52 ±0.31	1.05 ±0.28
Adler Grades I–III (Positive blood flow)	18	0.89 ±0.22	1.74 ±0.35
t		7.32	-7.16
P		< 0.001†	< 0.001†

Note: A total of 2 independent tests were performed in this table. The significance threshold after Bonferroni correction

Association with early imaging response: miR-205-5p expression in the favorable early-response group was significantly higher than that in the poor early-response group (1.19 ± 0.34 vs. 0.72 ± 0.28, mean difference = 0.47, 95% CI: 0.28–0.66, P < 0.001), while RUNX2 expression was significantly lower (1.36 ± 0.41 vs. 2.05 ± 0.38, mean difference = -0.69, 95% CI: -0.94 to -0.44, P < 0.001) (Fig. 9).

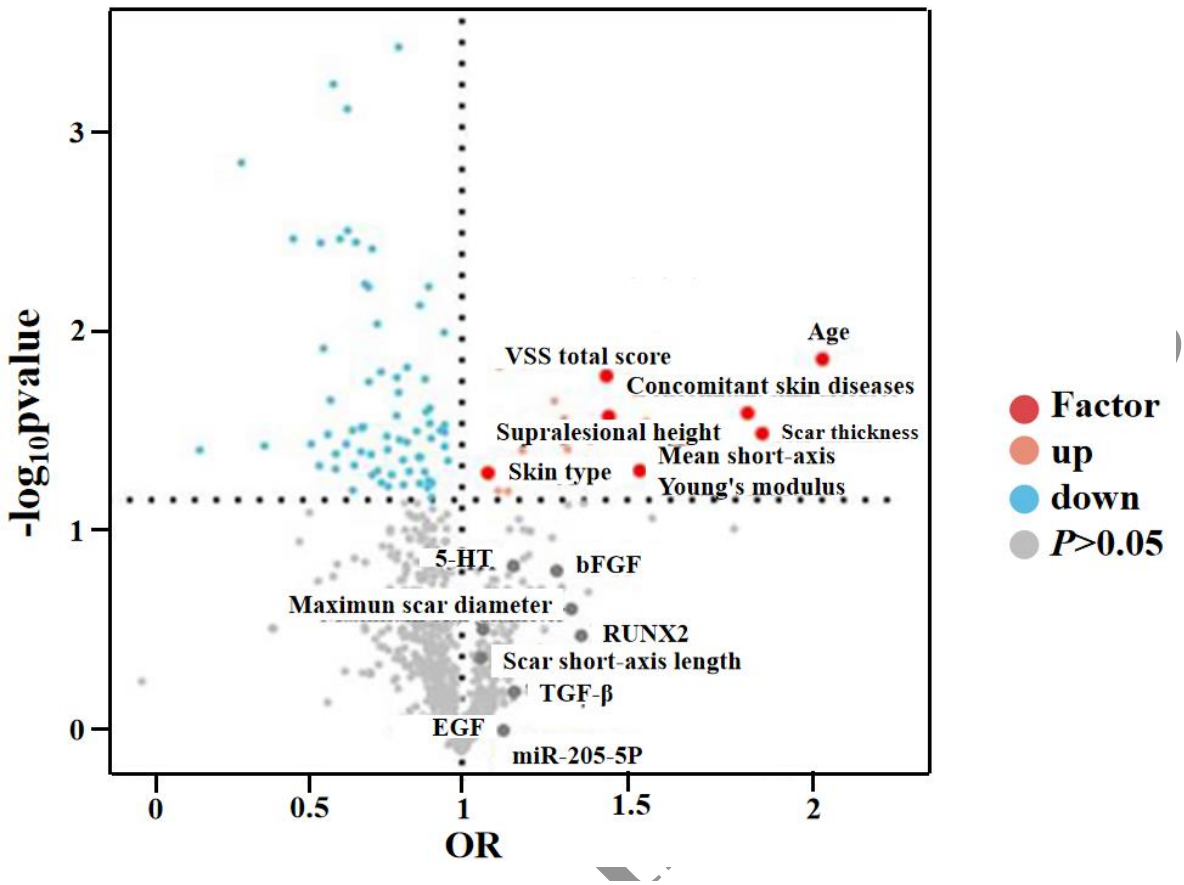


Fig. 10: Volcano plot of factors associated with poor early response (multivariable logistic regression analysis). The x-axis represents the odds ratio (OR) and the y-axis represents $-\log_{10}$(P value). Red dots indicate risk factors (OR > 1, P < 0.05), blue dots indicate protective factors (OR < 1, P < 0.05), and gray dots indicate non-significant variables (P > 0.05). Dashed vertical line, OR = 1; dashed horizontal line, significance threshold. n = 150.

Univariate and multivariable logistic regression analysis for poor early response

Univariate analysis identified the following variables associated with poor early response at $P < 0.10$: age, sex, disease duration, lesion location, etiology type, baseline VSS total score, scar thickness, scar short-axis length, supralesional height, short-axis Young's modulus mean, Adler Doppler vascularity grade, echo distribution and skin type. These variables were entered as candidates for the multivariable model. Before constructing the multivariable regression model, we evaluated the potential impact of treatment intensity variables (injection dose per unit area and total injection volume) on efficacy. Univariate analysis showed no significant differences between the favorable and poor early-response groups in injection dose per unit area (0.14 ± 0.04 mL/cm² in the favorable group vs. 0.15 ± 0.05 mL/cm² in the poor group, P = 0.342) or total injection volume (0.52 ± 0.23 mL vs. 0.56 ± 0.25 mL, P = 0.286). These dose-related variables were therefore not entered into the multivariable regression model.

Multivariable logistic regression analysis (Fig. 10) identified several independent risk factors for poor early response. In the final model, baseline scar thickness (OR = 2.41, 95% CI: 1.14–5.10, P = 0.021), Adler Doppler vascularity grade (grades I–III vs. grade 0: OR = 2.86, 95% CI: 1.35–6.06, P = 0.006), age (OR = 2.60, 95% CI: 1.18–5.74, P = 0.018), scar severity (OR = 4.01, 95% CI: 1.03–15.54, P = 0.045), baseline VSS total score (OR = 3.20, 95% CI: 1.14–8.95, P = 0.027), supralesional height (OR = 1.34, 95% CI: 1.02–1.76, P = 0.038), skin type (OR = 4.07, 95% CI: 1.06–15.67, P = 0.041) and short-axis Young's modulus mean (OR = 1.13, 95% CI: 1.01–1.26, P = 0.033) were independently associated with poor early response. A total of 8 predictors were tested in the final multivariable regression model. After Bonferroni correction, the significance threshold was adjusted to $\alpha' = 0.05/8 \approx 0.006$. After correction, Adler Doppler vascularity grade (corrected P = 0.048) and baseline scar thickness (corrected P = 0.168) were the two most prominent predictors; however, only Adler vascularity grade reached the corrected significance threshold, while the remaining variables with nominal P < 0.05 should be regarded as exploratory. Therefore, the statistical evidence for each predictor was considered exploratory, and its independent predictive value requires further confirmation in external validation.

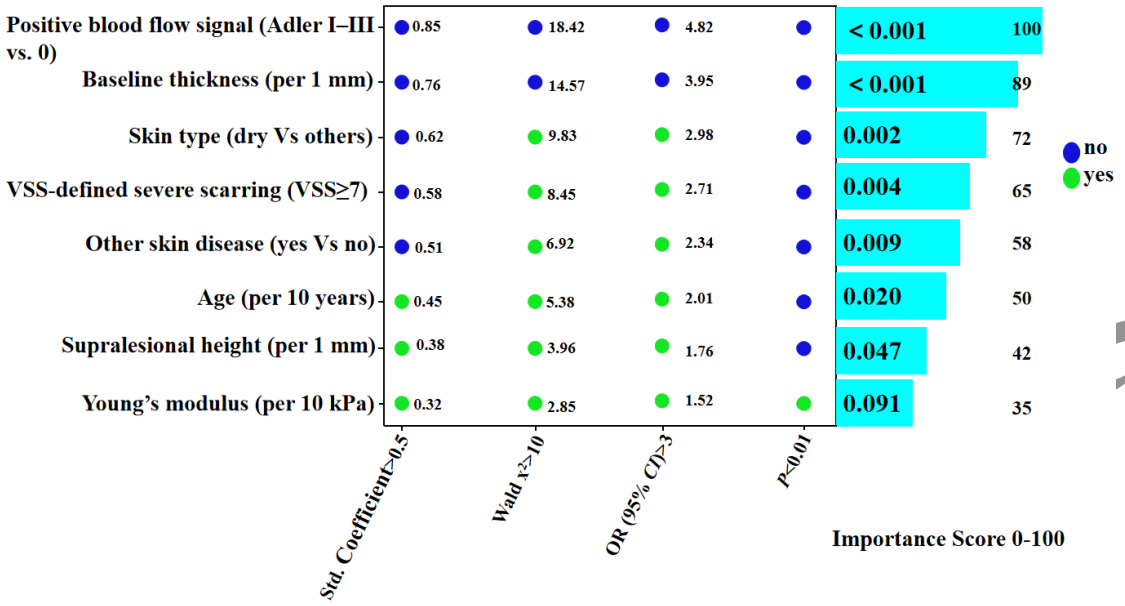


Fig. 11: Importance ranking of variables affecting poor early response (based on multivariable logistic regression). Standardized coefficient > 0.5 indicates strong influence; Wald χ^2 > 10 indicates high significance; OR 95% CI > 3 indicates a strong risk factor; P < 0.01 indicates extreme significance. The importance score (0–100) was calculated from the standardized coefficient, Wald χ^2, OR (95% CI) and P value. n = 150.

Variable importance ranking: Positive intralesional vascularity (100 points) and baseline scar thickness (89 points) emerged as the most significant predictors, substantially outperforming other variables such as skin type (72 points), scar severity (65 points) and age (50 points) (Fig. 11). Adler vascularity grade showed a significant inverse dose–response relationship with ΔTh%4W (Spearman ρ = −0.72, 95% CI: −0.81 to −0.62, P < 0.001), with each increment in vascularity grade associated with an average decrease of approximately 6–7% in ΔTh%4W.

Collinearity diagnostics showed VIF values ranging from 1.12 to 3.45 for all included variables, excluding multicollinearity concerns. Overall model performance: The combined HFUS model demonstrated an original AUC of 0.90 (95% CI: 0.85–0.95) for predicting poor early response, outperforming the clinical-variables-only model (AUC = 0.78, 95% CI: 0.72–0.85). The Hosmer–Lemeshow test indicated acceptable model calibration (P = 0.401). Model internal validation and calibration: After bootstrap internal validation (1,000 resampling iterations), the optimism-corrected AUC was 0.85 (95% CI: 0.80–0.91), representing a reduction of 0.04 from the original AUC (0.90), indicating mild overfitting but still acceptable overall discrimination. The calibration plot showed good agreement between predicted and observed probabilities (Fig. 12). The calibration slope was 0.89 (95% CI: 0.78–1.00) and the Brier score was 0.15 (range 0–0.25). A calibration slope < 1 is consistent with overfitting and aligns with the observed reduction in optimism-corrected AUC.

Ultrasound-based activity stratification system and efficacy outcomes

Based on baseline scar thickness (<5 mm, 5–7 mm, >7 mm) and Adler vascularity grade (grade 0 vs. grades I–III), keloids were classified into three ultrasound-based activity subtypes (Table 12): low-activity type (thickness <5 mm + Adler grade 0): favorable early response rate was 92.9% (26/28); intermediate-activity type (thickness 5–7 mm or Adler grade I): early response was uncertain, with comparable rates of favorable and poor responses; high-activity type (thickness >7 mm + Adler grades II–III): poor early response rate was 63.2% (12/19) (Fig. 13A). The low-activity type exhibited a molecular phenotype characterized by high miR-205-5p and low RUNX2 expression, while the high-activity type exhibited the opposite pattern (low miR-205-5p/high RUNX2), with statistically significant differences between groups (P < 0.001, Fig. 13B).

Week-24 recurrence and treatment durability

Of the 150 patients who completed the primary efficacy assessment at Week 12, 136 (90.7%) completed the Week-24 follow-up. Fourteen patients were lost to follow-up: 8 were unable to return due to work or residence relocation and 6 declined further participation for personal reasons. According to the pre-specified recurrence criteria (increased scar thickness, worsening clinical symptoms [pruritus/pain], or deterioration of VSS scores after initial improvement at Week 12), 28 patients (20.6% of those who completed follow-up) experienced recurrence by Week 24.

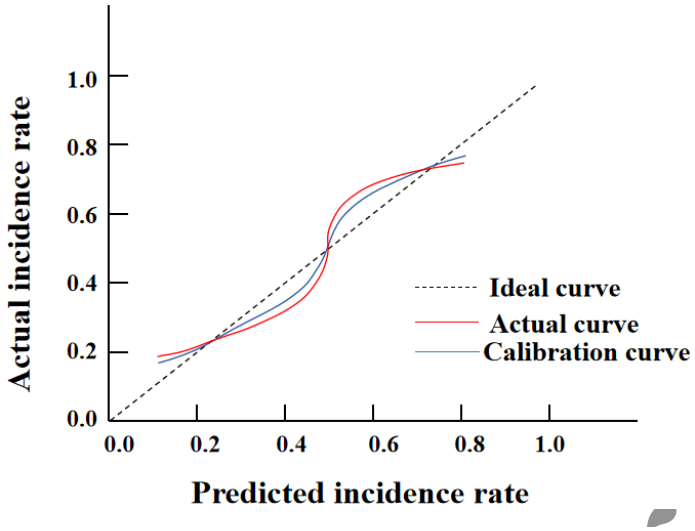


Fig. 12: Calibration curve of the multivariable logistic regression model. The x-axis represents the predicted probability of poor early response and the y-axis represents the observed proportion. The dashed diagonal line indicates ideal calibration (predicted = observed), the solid line represents the model calibration curve, and the gray shaded area represents the 95% CI. Calibration slope = 0.89; Brier score = 0.15; Hosmer‑Lemeshow test P = 0.401. Calibration was assessed after bootstrap internal validation (1,000 resampling iterations); n = 150.

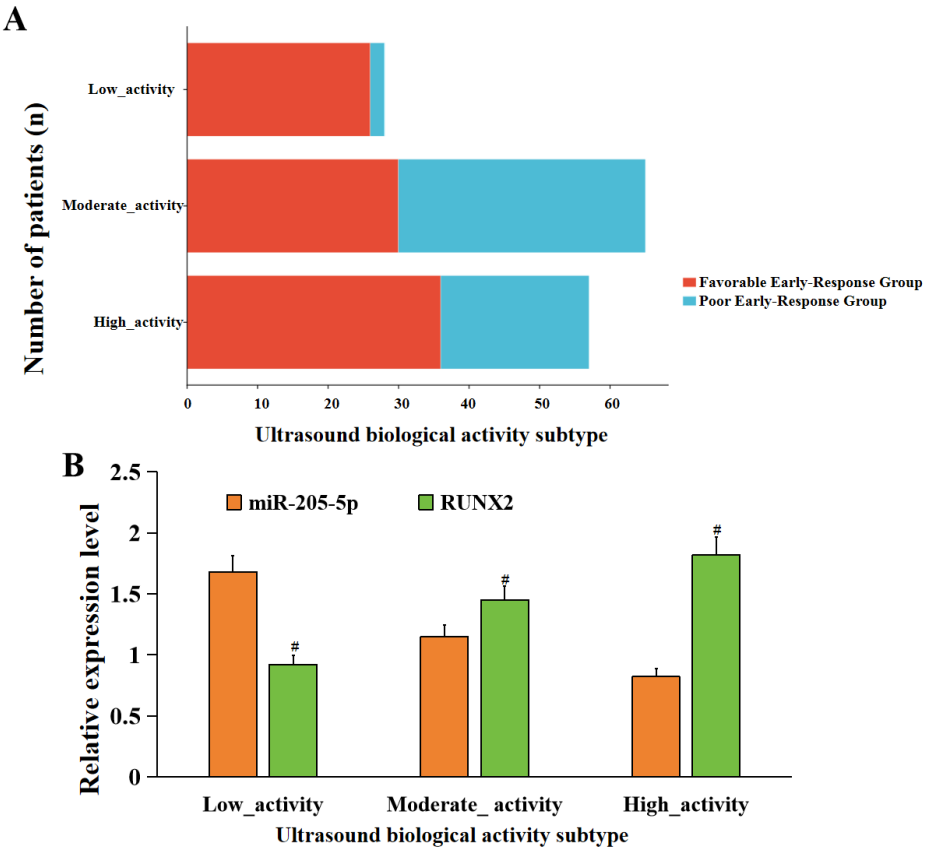


Fig. 13: Biological validation of the ultrasound-based activity stratification system. (A) Stacked bar chart of early imaging response distribution across different ultrasound activity types; (B) molecular marker expression levels in patients with different ultrasound activity types. Low-activity type, n = 10; intermediate-activity type, n = 18; high-activity type, n = 12. Data are shown as mean ± SD. *P < 0.05 vs. the favorable early-response group; #P < 0.05 vs. miR-205-5p (one-way ANOVA with post hoc test).

Table 13: Week-24 recurrence rates stratified by early imaging response grouping and ultrasound-based activity typing.

Stratification variable	Category	n	Recurrence rate [n (%)]	Risk difference (95% CI)	*P*-value
Early imaging response grouping	Favorable early-response (ΔTh%₍ ₄ W₎ ≥ 30%)	92	11 (12.0)	-	—
	Poor early-response (ΔTh%₍ ₄ W₎ < 30%)	58	17 (29.3)	17.3% (4.2%–30.5%)	0.006
Ultrasound-based activity typing	Low-activity type (thickness < 5 mm + Adler grade 0)	28	2 (7.1)	-	—
	Intermediate-activity type (thickness 5–7 mm OR Adler grade I)	103	16 (15.5)	8.4% (-7.8%–24.6%)	0.256
	High-activity type (thickness > 7 mm + Adler grades II–III)	19	10 (52.6)	45.5% (24.1%–66.9%)	<0.001

Note: A total of 2 independent χ^2 tests (early imaging response grouping and ultrasound-based activity typing) were

Stratified by early imaging response grouping (Table 13): among the favorable early-response group (n = 92), 11 patients (12.0%) experienced recurrence; among the poor early-response group (n = 58), 17 patients (29.3%) experienced recurrence, with a risk difference of 17.3% (95% CI: 4.2%–30.5%, P = 0.006).

Stratified by ultrasound-based activity typing (Table 13): 2 patients (7.1%) in the low-activity type (n = 28) experienced recurrence, 16 (15.5%) in the intermediate-activity type (n = 103) and 10 (52.6%) in the high-activity type (n = 19). The risk difference between the high-activity and low-activity types was 45.5% (95% CI: 24.1%–66.9%, P < 0.001), between the high-activity and intermediate-activity types was 37.1% (95% CI: 16.8%–57.4%, P < 0.001), while the difference between the low- and intermediate-activity types was not statistically significant (risk difference = 8.4%, 95% CI: -7.8%–24.6%, P = 0.256).

DISCUSSION

The progression and regression of keloid scars are essentially driven by their internal biological activity (Haque *et al*., 2025). Previous studies have shown that COL1A1 expression in lesional tissues is higher than in normal skin, while MMP1 activity is decreased, together leading to a serious imbalance between collagen synthesis and degradation (Roh *et al*., 2025). p-AKT (Thr308 site) expression is 2.85 times higher than in normal control samples and intervention targeting the PI3K/AKT/mTOR pathway can reduce collagen synthesis by 58.70% (Liu *et al*., 2025). These highly metabolic, actively proliferating lesions often manifest as increased thickness, internal hypoechoic signals, and abundant intralesional vascularity on ultrasound (Chen et al., 2021), and collagen fiber alignment entropy in active keloids (0.72±0.08) is significantly higher than in mature lesions (Meng *et al*., 2023).

A quantitative relationship between baseline HFUS parameters (scar thickness and blood flow grading) and early imaging response was systematically characterized. For every 1 mm increase in baseline scar thickness, the risk of poor early response increased approximately 2.4-fold (OR=2.412). Collagen fibers within keloids form a dense, highly cross-linked network with porosity significantly lower than that of the normal dermis (Mardhiah *et al*., 2025). When scar thickness exceeds 5 mm, the effective concentration of glucocorticoid molecules (molecular weight 434.5 Da) penetrating the lesion core via simple diffusion decreases exponentially (Ardehali*et al*., 2007; Branyiczky *et al*., 2025).

ΔTh%□□W□ in the thickness >7 mm group was only 15.6% (blood flow-positive) to 24.1% (no blood flow), significantly lower than the 38.2% observed in the <5 mm group, providing indirect support for the "thickness-permeability barrier" hypothesis. Keloid fibroblasts show significant spatial heterogeneity in proliferative activity across lesion depths (Komulainen *et al*., 2024). Fibroblasts in the deep lesion area near the subcutaneous fat layer exhibit higher α-SMA expression and collagen synthesis rates (Roh *et al*., 2025), alongside reduced glucocorticoid sensitivity. As scar thickness increases, the absolute population of these deep-seated, drug-resistant fibroblasts expands (Lobos *et al*., 2017), making it difficult to suppress their activity even with partial drug penetration. This is supported by the observation that post-treatment bFGF levels in the poor early-response group remained significantly higher than those in the favorable group, suggesting that residual activated fibroblasts in deep tissues continue to secrete pro-fibrotic factors.

In the multivariable model, intralesional vascularity (100 points) was more predictive than scar thickness (89 points), and Adler Doppler vascularity grade showed a clear inverse dose-response relationship with ΔTh%□W. The abundance of intralesional vascularity directly reflects the microvascular density and perfusion status inside keloid scars (Tang *et al*., 2025b). Neovascularization not only provides oxygen and nutrients to fibroblasts (Ningsih *et al*., 2024) but also secretes multiple pro-fibrotic factors,

facilitating a "vascular-fibroblast" positive feedback loop (Sun *et al*., 2024). Studies have found that the Piezo1 mechanosensitive channel is expressed at higher levels in keloids and its activation can enhance fibroblast proliferation (Liu *et al*., 2025). Blood flow shear stress activates Piezo1, raising the possibility that vascular-rich lesions may exhibit enhanced fibroblast activity through mechanotransduction. Although the anti-inflammatory efficacy of glucocorticoids partially depends on inducing microvascular constriction and inhibiting vascular permeability (Ningsih *et al*., 2024), their impact is markedly limited once extensive Adler Grade II--III neovascularization networks are established (Cho *et al*., 2024). Lesions with abundant vascularity receive continuous inflammatory signaling input, partially counteracting glucocorticoid inhibition.

In interpreting these associations, several baseline characteristics differed significantly between the favorable and poor early-response groups. Patients in the poor early-response group were older, had a higher proportion of dry skin, higher baseline VSS scores, more lesions ≥4 cm and more concomitant skin diseases. These factors may independently influence treatment response — aging reduces skin repair capacity, dry skin may reflect impaired barrier function, and higher VSS scores indicate more severe scar burden (Knowles and Glass, 2023). These baseline differences may partially confound the observed association between ultrasound parameters and efficacy. To address this, age, baseline VSS score, skin type, lesion size and comorbidities were included as candidate variables in the multivariable model. After adjustment, scar thickness and vascularity grade remained independently associated with poor early response, suggesting their predictive value extends beyond known clinical confounders. Nevertheless, residual and unmeasured confounders (e.g., Fitzpatrick skin phototype, previous scar behavior) may still exist. The interplay between clinical features and ultrasound parameters warrants further consideration: the imaging phenotype of increased thickness and rich vascularity may serve as an integrated "imaging proxy" for a more severe clinical baseline, rather than a purely independent biological driver. Although VIF diagnostics excluded multicollinearity, the intrinsic association between clinical risk factors and imaging parameters cannot be fully disentangled, and this residual covariance may modestly inflate the apparent predictive weight of ultrasound parameters. Readers should therefore interpret the predictive importance of thickness and vascularity within the context of the overall clinical picture, recognizing that ultrasound findings and clinical risk factors are complementary rather than fully independent.

A robust correlation was also identified between baseline HFUS parameters (increased thickness and positive blood flow) and a molecular profile characterized by low miR-205-5p and high RUNX2 expression—a phenotype significantly more pronounced in the poor early-response group. These molecular findings should be interpreted as exploratory biological validation rather than definitive mechanistic evidence. miR-205-5p serves as a critical negative regulator of fibrotic progression; under physiological conditions, it binds to the 3'-UTR of TGFBR2, inhibiting its translation and limiting excessive TGF-β signaling activation. miR-205-5p expression in the poor early-response group decreased by approximately 40% compared with the favorable group, a decrease sufficient to cause 2- 3-fold upregulation of TGFBR2 expression. Excessive TGF-β signaling directly promotes transcription of collagen genes such as COL1A1 and COL3A1 (Roh *et al*., 2025) and enhances fibroblast-to-myofibroblast differentiation through Smad3 phosphorylation (Prananda *et al*., 2025), exerting stronger contractile and secretory capabilities. RUNX2, recognized as the master transcription factor for osteogenic differentiation and tissue calcification (Kong *et al*., 2025), has recently been confirmed to be abnormally upregulated in keloids. RUNX2 expression showed strong positive correlations with baseline scar thickness ($r = 0.921$) and blood flow grading ($r = 0.625$). This likely reflects RUNX2 binding to specific elements within the COL1A1 promoter (Hamzehlou et al., 2026) to activate collagen gene transcription; it also forms transcription complexes with Smad3, synergistically enhancing TGF-β pro-fibrotic effects (Tang *et al*., 2025a). The significantly higher RUNX2 expression in the poor early-response group suggests RUNX2 overexpression may serve as a pivotal "fibrosis-associated molecular phenotype" contributing to therapy resistance. A potential negative regulatory axis exists between miR-205-5p and RUNX2: bioinformatics predictions identified a conserved miR-205-5p binding site within the 3'-UTR of RUNX2 (Zhang *et al*., 2020; Mo *et al*., 2021), suggesting that miR-205-5p downregulation relieves inhibitory constraint on RUNX2. The two expression levels showed a significant negative correlation ($r=-0.682$), supporting this hypothesis.

These exploratory molecular findings support the proposed ultrasound-based activity stratification. The low-activity type (thickness <5 mm, no blood flow) exhibited a protective molecular phenotype characterized by high miR-205-5p and low RUNX2 expression, with fibroblasts in a relatively quiescent state; glucocorticoids can effectively maintain this inhibited state, resulting in a favorable efficacy rate of 92.9%. The high-activity type (thickness >7 mm, positive blood flow) showed a pathogenic molecular phenotype characterized by low miR-205-5p and high RUNX2 expression, reflecting a biologically active fibrotic phenotype that is difficult to reverse with injection monotherapy and yielded a poor efficacy rate of 63.2%. Based on these exploratory analyses, the proposed ultrasound-based activity stratification system suggests the following hypotheses for future testing rather than clinical recommendations:

Article in Press

patients with the low-activity subtype may achieve satisfactory efficacy with injection monotherapy, but this requires prospective confirmation; patients with the high-activity subtype may benefit from more intensive strategies such as early combination with pulsed dye laser or 5-fluorouracil, but this requires formal prospective testing; for lesions with extreme thickness (≥10 mm), surgical excision followed by adjuvant radiotherapy could be considered in theory, but the number of such lesions was limited and feasibility requires further evaluation. All therapeutic suggestions are derived from exploratory analyses and should not be interpreted as clinical recommendations; they require validation in prospective, multi-center cohorts before formal clinical guidance can be formulated.

The selection of Week-4 ultrasound thickness change rate rather than long-term recurrence as the primary endpoint reflects a clinically demand-driven design. In clinical practice, Week 4 falls within the inter-injection interval: it is a reasonable time point for assessing initial response and an opportunity to modify strategy before the second injection. Ultrasound response at this time point predicted Week-12 overall efficacy (AUC=0.897, positive predictive value 83.7%), meaning clinicians can obtain sufficient information at 4 weeks after the first injection to decide whether to maintain, intensify, or switch treatment strategies. Thus, ΔTh%□W serves not merely as a research endpoint but as a clinical decision-support tool.

Compared with previous HFUS studies, several contributions are offered. One study (Lobos *et al*., 2017) found that lesions with thickness >5 mm had significantly higher proportions of positive vascularity, consistent with the present findings; however, another study reported a weaker correlation between perfusion parameters and VSS (ρ = 0.308) than that between the VSS thickness subscale and the ultrasound-derived thickness change rate (r = 0.814), suggesting that objective ultrasound measurements may be more sensitive to subtle changes For treatment monitoring, one study (Cai *et al*., 2020) used high-frequency ultrasound to monitor injection efficacy longitudinally and reported a mean reduction in keloid dermal thickness of approximately 39.0%, but it focused on dynamic changes during treatment rather than baseline prediction. By shifting the analytical window forward to baseline, evidence is provided that a single pre-treatment HFUS measurement can predict Week-4 early response. Furthermore, whereas another study (Cao *et al*., 2024) reported a significant correlation between SWE-derived Young's modulus and histological collagen density, SWE was further shown to independently predict early imaging response (OR = 1.128), extending its potential from a static marker to a dynamic predictor of therapeutic response Notably, even in the relatively mature field of laser therapy for scars, significant discrepancies in reported efficacy exist across studies---a systematic review demonstrated that some studies reported fractional CO□ laser achieving 49.8% VSS reduction, significantly superior to Er:YAG laser (28.2%), while others found no significant difference at three months (Haji *et al*., 2025). This inter-study inconsistency suggests heterogeneity in treatment response exists not only across patients but also across study designs and evaluation criteria, further reinforcing the need for a unified, objective efficacy prediction tool such as the proposed ultrasound-based stratification system.

From a personalized treatment decision-making perspective, HFUS adds value not only by predicting short-term efficacy but also by providing quantifiable risk stratification before treatment initiation. Traditionally, intralesional injection therapy for keloids has followed a "one-size-fits-all" approach---uniform dosage, interval and course---with adjustments made only passively when poor efficacy becomes apparent. This reactive strategy delays effective treatment and incurs unnecessary drug exposure and economic burden. Through baseline HFUS assessment, clinicians can classify patients into low-, intermediate-, or high-activity types before the first injection, enabling preemptive differentiation of treatment pathways: low-activity patients can safely receive standard protocols without overtreatment, while high-activity patients may benefit from adjunctive therapy at the first session or a shorter evaluation window (e.g., 2--3 weeks) to facilitate timely strategy adjustment. This "predict-stratify-match" paradigm shifts treatment decisions from reactive to proactive, representing the most significant practical value of HFUS as a precision medicine tool in keloid management.

It is acknowledged that the sample size calculation was based on a two-group comparison of the primary outcome rather than strictly on prediction model development requirements. This approach is commonly adopted in the preliminary phase of exploratory prediction research; however, the predictive model findings from this study are exploratory and require external validation in an independent prospective cohort before generalization. Based on the number of poor early-response events (58 cases) and the number of variables in the final model (8), the calculated EPV was 58/8 = 7.25, which is below the commonly recommended minimum of EPV ≥ 10, suggesting a risk of overfitting. To mitigate this, conservative variable selection (entry criterion $P < 0.05$), bootstrap internal validation (1,000 resampling iterations), and restriction of the final model to 8 variables were used. Although stepwise regression determined the final model composition, candidate predictors were selected based on pre-specified clinical and biological plausibility criteria established during protocol development, rather than selected post hoc after data exploration. Sensitivity analysis results are provided as supplementary references, while the main conclusions are based on uncorrected

In Press

nominal P values. Given the exploratory nature of this study, findings with P values between 0.01 and 0.05 were considered hypothesis-generating and requiring further validation in independent cohorts.

Several additional limitations should be acknowledged. First, this was a single-center study with a predominantly East Asian cohort (92.0% Fitzpatrick type III–IV) and the eligibility criteria were relatively restrictive (only localized lesions with longest diameter ≤ 10 cm, located on the trunk or extremities and not belonging to the severe subgroup with extreme tension abnormalities or significant joint functional impairment requiring surgical reconstruction as first-line treatment), limiting generalizability. The findings apply specifically to untreated, localized, adult truncal/extremity keloids treated with injection monotherapy and caution is needed when extrapolating to pediatric patients, giant lesions, previously treated/recurrent lesions, or lesions with extreme tension abnormalities or significant joint functional impairment. Second, both the predictive model and ultrasound-based stratification system were developed and internally validated, and the optimism-corrected AUC was 4.8% lower than the original estimate, indicating a moderate overfitting risk; external validation in independent prospective cohorts is mandatory. Third, although ultrasound measurements demonstrated good inter-observer agreement (ICC 0.91–0.95), they are inherently operator-dependent; measurement accuracy may decline without rigorous training and quality control. Fourth, the molecular analyses were based on an exploratory subgroup of only 40 patients, with small samples in the high-activity type (n=12) and blood flow-positive group (n=18); tissue samples were obtained from surgical excisions rather than random biopsies, representing a subgroup with greater disease burden (high-activity type 35.0% vs. 12.7% in the overall cohort) and small samples tend to overestimate effect sizes—the true effect size of observed strong correlations may be smaller and whether the observed molecular patterns represent the broader keloid population remains unclear. Fifth, recurrence follow-up was limited to 24 weeks, which may underestimate long-term cumulative recurrence rates. Sixth, several clinically relevant confounders, including Fitzpatrick skin phototype, detailed lesion morphology (nodular vs. flat), anatomical tension, and prior scar behavior (e.g., recurrence history), were not systematically collected; these unmeasured factors may confound the association between ultrasound parameters and treatment response. Additional limitations include the absence of real-time ultrasound guidance during injection, unclear applicability to triamcinolone-based regimens, potential false-positive risk from multiple hypothesis testing, and limited access to HFUS equipment and trained operators.

Despite these limitations, this study provides a valuable exploratory framework for imaging-based prediction of corticosteroid response in keloids. Future research should conduct multi-center prospective cohorts for external validation, employ random biopsies or prospective molecular cohort designs to avoid selection bias, extend follow-up duration to evaluate long-term recurrence, incorporate Fitzpatrick skin phototype and standardized lesion morphology and tension measurements for comprehensive confounder adjustment, develop automated ultrasound image analysis tools to reduce operator dependence and perform functional experiments to validate the biological significance of the miR-205-5p/RUNX2 regulatory axis.

CONCLUSION

In summary, baseline HFUS thickness and Doppler vascularity showed statistical associations with early imaging response, suggesting their potential as non-invasive indicators for predicting early imaging response (Week 4) and primary efficacy (Week 12) to intralesional corticosteroid therapy in keloids. However, these associations are observational and do not establish a causal relationship for the predictive model. The correlation between high-activity ultrasound features and the miR-205-5p/RUNX2 expression pattern is exploratory and based on correlational analysis; no regulatory relationship or causation can be inferred and its mechanistic significance requires functional validation. Exploratory Week-24 follow-up showed that the recurrence rate in the high-activity type (52.6%) was higher than that in the low-activity type (7.1%), suggesting a possible association between baseline ultrasound typing and long-term recurrence risk; however, this finding requires confirmation in larger cohorts. Furthermore, the correlation between high-activity ultrasound features and the miR-205-5p/RUNX2 expression pattern is preliminary, limited by the small sample size of the molecular subgroup and its effect size estimate may be unstable; its mechanistic significance requires validation in larger samples.

Acknowledgment
The necessary equipment for this study was gratefully provided by the Affiliated Hospital of Nantong University.

Authors' contribution
Tianyi Xie: Performed the majority of the data acquisition and analysis, including statistical modeling and participated in manuscript writing.
Xiaodong Chen: Conceptualized the research, designed the study methodology, supervised the entire research project and reviewed and finalized the manuscript.
All authors participated in data interpretation and manuscript revision and agreed to submit and publish the final version.

Funding
This research received no external funding.

Article in Press

Data availability statement

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Ethical approval

This study was approved by the Ethics Committee of the Affiliated Hospital of Nantong University (Approval No. 2025-Q240-04). This study was performed in adherence with the STROBE guidelines. See supplementary file for the STROBE checklist.

Conflict of interest

There are no conflicts to declare.

REFERENCES

Adler DD, Carson PL, Rubin JM and Quinn-Reid D (1990). Doppler ultrasound color flow imaging in the study of breast cancer: Preliminary findings. *Ultrasound Med Biol*., **16**(6): 553-559.

Ardehali B, Nouraei SA, Van Dam H, Dex E, Wood S and Nduka C (2007). Objective assessment of keloid scars with three-dimensional imaging: Quantifying response to intralesional steroid therapy. *Plast Reconstr Surg*., **119**(2):556-561.

Baryza MJ and Baryza GA (1995). The Vancouver Scar Scale: an administration tool and its interrater reliability. *J Burn Care Rehabil*., **16**(5): 535-538.

Branyiczky MK, Metko D, Hsu JTS and Li MK (2025). Combination of carbon dioxide laser and triamcinolone acetonide for keloid and hypertrophic scar treatment: A systematic review. *J Am Acad Dermatol*., **93**(3): 821-825.

Cai L, Hu M, Lin L, Zheng T, Liu J and Li Z (2020). Evaluation of the efficacy of triamcinolone acetonide in the treatment of keloids by high-frequency ultrasound. *Skin Res Technol*., **26**(4): 489-493.

Cao LL, Yang ZG, Qi WH, Zhang H, Bi Y, Shan Y, Cui XW and Jiang F. (2024). A preliminary study on ultrasound techniques applied to evaluate the curative effect of botulinum toxin type a in hypertrophic scars. *Heliyon*., **10**(15): e34723.

Chen C, Liu S, Zhao C, Wang R, Yu N, Long X, Wang Y, Yang F, Sun J, Lu ZL, Xia Y, Jiang Y and Yang M (2021). Activity of keloids evaluated by multimodal photoacoustic/ultrasonic imaging system. *Photoacoustics*., **24**: 100302.

Cho H, Dohi T, Wakai H, Quong WL, Linh NDT, Usami S and Ogawa R (2024). In the face and neck, keloid scar distribution is related to skin thickness and stiffness changes associated with movement. *Wound Repair Regen*., **32**(4): 419-428.

Gold MH, Berman B, Clementoni MT, Gauglitz GG, Nahai F and Murcia C (2014). Updated international clinical recommendations on scar management: Part 1--evaluating the evidence. *Dermatol Surg*., **40**(8): 817-824.

Haji Mohammadi A, Seirafianpour F, Khosravi M, Jafarzadeh A, Neshastesaz Kashi H, Baradaran H and Goodarzi A (2025). A systematic review of comparative clinical trials on the efficacy, safety and patient satisfaction of ablative and non-ablative laser therapies for atrophic, hypertrophic and keloid scars. *Lasers Med Sci*., **40**(1): 280.

Hamzehlou S, Vishlaghi N, Shelton JM, Levi B, Xing C and Glass DA 2nd (2026). RUNX2 regulates aberrant expression of collagen genes in keloid fibroblasts. *J. Invest Dermatol*., **146**(5): 1403-1408.e1.

Haque R, Chang SE and Moon IJ (2025). Oxidative stress-mediated modulation of fibrosis and inflammation in keloid fibroblasts by cold atmospheric plasma. *Skin Pharmacol Physiol*., **38**(5-6): 185-197.

Knowles A and Glass DA 2nd (2023). Keloids and Hypertrophic Scars. *Dermatol Clin*., **41**(3): 509-517.

Komulainen T, Daymond P, Hietanen KE, Kaartinen IS and Järvinen TAH (2024). Myofibroblasts reside in the middle dermis of the keloids but do not predict the response to injection therapies: A double-blinded, randomized, controlled trial. *Front Med (Lausanne)*., **11**:1293028.

Kong J, Zhou C, Qin H, Li C, Wu Z and Zhang L (2025). Identifying key genes, miRNAs and pathways in keloid formation: A bioinformatics and experimental study. *J Plast Reconstr Aesthet Surg*., **102**: 313-322.

Leelawattanachai J, Sittisaksomjai S and Disphanurat W (2025). Comparative effectiveness of 5-fluorouracil dissolving microneedle patch vs. 5-fluorouracil intralesional injection for the treatment of keloid scars: A randomised, single-blinded, split-scar study. *Wound Repair Regen*., **33**(4): e70078.

Lobos N, Wortsman X, Valenzuela F and Alonso F (2017). Color doppler ultrasound assessment of activity in keloids. *Dermatol Surg*., **43**(6): 817-825.

Liu J, Yan J, Qi S, Zhang J, Zhang X and Zhao Y (2025). The PI3K/AKT/mTOR pathway in scar remodeling and keloid formation: Mechanisms and therapeutic perspectives. *Front Pharmacol*., **16**: 1678953.

Mardhiah J, Halim AS, Heng S and Saipolamin AQ (2025). Scoping review for pain mitigation during intralesional injections of corticosteroid for hypertrophic scar and keloid treatment. *BMJ Open*., **15**(2): e092800.

Meng J, Wang G, Zhou L, Jiang S, Qian S, Chen L, Wang C, Jiang R, Yang C, Niu B, Liu Y, Ding Z, Zhuo S and Liu Z (2023). Mapping variation of extracellular matrix in human keloid scar by label-free multiphoton imaging and machine learning. *J Biomed Opt*., **28**(4): 045001.

Merlino L, Dominoni M, Pano MR, Pasquali MF, Senatori R, Zino G and Gardella B (2025). Recent progress in keloid mechanism and treatment: A comprehensive review. *Biomedicines*., **13**(9): 2276.

Murakami T and Shigeki S (2024). Pharmacotherapy for keloids and hypertrophic scars. *Int J Mol Sci*., **25**(9): 4674.

Article in Press

Mo C, Huang B, Zhuang J, Jiang S, Guo S and Mao X (2021). LncRNA nuclear-enriched abundant transcript 1 shuttled by prostate cancer cells-secreted exosomes initiates osteoblastic phenotypes in the bone metastatic microenvironment via miR-205-5p/runt-related transcription factor 2/splicing factor proline- and glutamine-rich/polypyrimidine tract-binding protein 2 axis. *Clin Transl Med*., **11**(8): e493.

Ningsih SS, Jusman SWA, Syaidah R, Nauli R and Fadilah F (2024). Efficient protocol for isolating human fibroblast from primary skin cell cultures: Application to keloid, hypertrophic scar and normal skin biopsies. *Biol Methods Protoc*., **9**(1): bpae082.

Prananda AT, Nugraha SE, Situmorang PC and Syahputra RA (2025). Comparative effectiveness of plant-derived compounds in keloid management: A review. *Front Pharmacol*., **16**: 1576851.

Roh H, Kim YH, Heo KJ, Hong JW and Lee WJ (2025). Antifibrotic effects of an α7 nicotinic acetylcholine receptor agonist in keloid fibroblasts and a rat scar model. *Int J Mol Sci*., **26**(18): 8868.

Sun H, Shen W and Zhong HJ (2024). The crucial role of surface roughness in scar keloid assessment. *Burns*., **50**(2): 535-536.

Tang J, Liu Q, Yang B and Lu H (2025a). Exosome encapsulation of miR-205-5p suppresses neuroblastoma progression by targeting RUNX2. W*orld J Pediatr Surg*., **8**(3): e000993.

Tang J, Huang Z and Yu P (2025b). The role of JPX in regulating FUS/SLC7A11 signaling pathway mediated ferroptosis in keloid fibroblasts and its potential in scar repair. *Biochem Biophys Res Commun*., **763**: 151770.

Wen J, Li Z, Li S, Jin L, Gao H, Yu N and Wang X (2026). Insights from comparing blood perfusion metrics and Vancouver Scar Scale in keloid assessment. *Burns*., **52**(1): 107766.

Willenborg S, Schönborn K, Sawant M, Bornikoel A, Yamane T, Zeinert I, Eckes B, Eming SA and Krieg T (2025). Fibroblast-derived TGFβ1 regulates skin repair and fibrosis. *Wound Repair Regen*., **33**(4): e70065.

Xu Q, Bao C, Liu J, Zou Y, Zhu Y, Shen S, Fang F, Wang S, Lin M, Huang J, Gong T, Cheng B, Huang Z and Ji C (2024). Effect of epidermoid cysts on the efficacy of intralesional corticosteroid therapy for hypertrophic scars and keloids: A prospective pilot study. *Dermatol Surg*., **50**(2): 160-164.

Zhang L, Liu L, Xu X, He X, Wang G, Fan C, Zheng Q and Li F (2020). miR-205/RunX2 axis negatively regulates CD44/CD24 breast cancer stem cell activity. *Am J Cancer Res*., **10**(6):1871-1887.

Article in press