3D ULTRASOUND BIOMICROSCOPY FOR ASSESSMENT OF CARTILAGE REPAIR TISSUE: VOLUMETRIC CHARACTERISATION AND CORRELATION TO ESTABLISHED CLASSIFICATION SYSTEMS

M. Schöne[1], N. Männicke[1], J.S. Somerson[2], B. Marquaß[3,4], R. Henkelmann[3,4,] T. Aigner[5], K. Raum[1] and R.M. Schulz[3,4,6,*]

[1]Berlin-Brandenburg School for Regenerative Therapies, Charité – Universitätsmedizin Berlin, Berlin, Germany
[2]Department of Orthopaedics and Sports Medicine, University of Washington, Seattle, USA
[3]Clinic of Orthopaedics, Trauma and Plastic Surgery, University Hospital Leipzig, Leipzig, Germany
[4]Translational Centre for Regenerative Medicine, University of Leipzig, Leipzig, Germany
[5]Institute of Pathology, Medical Centre Coburg, Coburg, Germany
[6]Department of Cell Techniques and Applied Stem Cell Biology, Centre for Biotechnology and Biomedicine, University of Leipzig, Leipzig, Germany

Abstract

Objective and sensitive assessment of cartilage repair outcomes lacks suitable methods. This study investigated the feasibility of 3D ultrasound biomicroscopy (UBM) to quantify cartilage repair outcomes volumetrically and their correlation with established classification systems. 32 sheep underwent bilateral treatment of a focal cartilage defect. One or two years post-operatively the repair outcomes were assessed and scored macroscopically (Outerbridge, ICRS-CRA), by magnetic resonance imaging (MRI, MOCART), and histopathology (O'Driscoll, ICRS-I and ICRS-II). The UBM data were acquired after MRI and used to reconstruct the shape of the initial cartilage layer, enabling the estimation of the initial cartilage thickness and defect volume as well as volumetric parameters for defect filling, repair tissue, bone loss and bone overgrowth. The quantification of the repair outcomes revealed high variations in the initial thickness of the cartilage layer, indicating the need for cartilage thickness estimation before creating a defect. Furthermore, highly significant correlations were found for the defect filling estimated from UBM to the established classification systems. 3D visualisation of the repair regions showed highly variable morphology within single samples. This raises the question as to whether macroscopic, MRI and histopathological scoring provide sufficient reliability. The biases of the individual methods will be discussed within this context. UBM was shown to be a feasible tool to evaluate cartilage repair outcomes, whereby the most important objective parameter is the defect filling. Translation of UBM into arthroscopic or transcutaneous ultrasound examinations would allow non-destructive and objective follow-up of individual patients and better comparison between the results of clinical trials.

Keywords: Cartilage – repair/regeneration, ultrasound biomicroscopy, Outerbridge, ICRS-CRA, MOCART, O'Driscoll, ICRS-I, ICRS-II.

*Address for correspondence:
Dr. Ronny Schulz
University of Leipzig,
Centre for Biotechnology and Biomedicine,
Deutscher Platz 5, 04103 Leipzig, Germany
Telephone number: +49 341 9731352
Fax number: +49 341 9731359
Email: ronny.schulz@bbz.uni-leipzig.de

Introduction

In recent years, the spectrum of cartilage repair techniques has increased remarkably. Macroscopic evaluation, magnetic resonance imaging (MRI) and histological preparation are usually used to characterise the cartilage repair outcomes. For each method several scoring systems were developed, which have been partially validated and recommended by the International Cartilage Repair Society (ICRS) (Buschmann and Saris, 2011).

The depth of native articular cartilage lesions can be scored macroscopically and intraoperatively by the surgeon according to the Outerbridge score (Outerbridge, 1961) or by the newer and more specific ICRS Grade. For cartilage repair outcomes the ICRS recommends the Cartilage Repair Assessment (CRA) scoring system in the ICRS Cartilage Injury Evaluation Package with three distinct categories: "degree of defect repair", "integration to border zone", and "macroscopic appearance". However, intraoperative scoring according to Outerbridge, ICRS, and ICRS-CRA scales involves a large subjective component resulting in high inter-observer variability (Bonasia *et al*., 2015; Cameron *et al*., 2003). In addition, macroscopic judgement of the cartilage repair outcome does not allow for assessment of the cartilage thickness unless the bone is exposed. Furthermore, the repair tissue can mask problems in the sub-superficial repair region; an apparently well-filled lesion may hide undesirable malformation of the subchondral bone plate. However, in recent years, the awareness of the subchondral bone as one important aspect of cartilage repair has increased (Gomoll *et al*., 2010).

MR imaging provides visualisation of the whole joint with high soft tissue contrast and allows for assessment of the cartilage layer, the underlying bone as well as surrounding soft tissue. It can also be used for non-invasive follow-up of cartilage repair. According to the recommendations of the ICRS, assessment of the cartilage layer and the underlying bone compartment should be performed according to the nine categories of the 2D Magnetic Resonance Observation of Cartilage Repair Tissue (MOCART) scheme (Ebert *et al*., 2014; Marlovits *et al*., 2006). As already shown, some MRI-based classification systems and single items of the MOCART score correlate with clinical outcomes after cartilage repair (Blackman *et al*., 2013; Windt *et al.,* 2013a). However, the correlation strength also changes depending on the repair approach. At this time, "no current MRI classification system has been shown to correlate with clinical outcomes

after all types of cartilage repair surgery" (Blackman *et al.*, 2013) and "there remains a critical lack of conclusive evidence to determine whether morphological MRI is reliable in predicting clinical outcomes after cartilage repair" (Windt *et al.,* 2013b).

For histopathological grading, a wide variety of systems have been developed to describe the quality of the cartilage repair tissue (Rutgers *et al.*, 2010). Some scoring systems, *e.g.* O'Driscoll (O'Driscoll *et al.*, 1988), were developed for animal studies, where the whole joint is available for examination. In contrast, the scoring systems like the ICRS-I (Mainil-Varlet *et al.*, 2003) and ICRS-II (Mainil-Varlet *et al.*, 2010) were developed for the evaluation of human cartilage repair tissue, where only a 2 mm diameter biopsy from the central defect site is available. Except for the ICRS-II score, all of these histopathological grading systems use an ordinal scale to describe the *in vivo* repair result based on cross-sections of the cartilage-bone sample.

The inherent problem of all macroscopic, MRI, and histopathological cartilage scoring systems is that they are prone to typical systematic errors of interventional and observational studies: *i.e.*, detection bias, information bias, selection bias, and observer bias. Therefore, the U.S. Food and Drug Administration (Web ref. 1) and the Committee for Advanced Therapies (CAT) of the European Medicines Association (EMA) (Web ref. 2) recommended patient-based outcome measures such as the Knee injury and Osteoarthritis Outcome Score (KOOS) or according to the International Knee Documentation Committee (IKDC) (Irrgang *et al.*, 2001; Roos *et al.,* 1998) as primary or co-primary endpoints in pivotal clinical trials, since there is a "current lack of other outcome measures that are both sensitive and objective" (Web ref. 1).

Cartilage repair has been investigated for decades. However, until now no overall satisfactory approach has been found and the results are a "long way from achieving our goal of cartilage regeneration" (Hunziker *et al.*, 2015). Therefore, the need for "a more thorough and systematic approach" to the evaluation of cartilage repair was recently identified (Hunziker *et al.,* 2015).

In this study we evaluated the feasibility of ultrasound biomicroscopy (UBM) as an objective method for morphometrical analysis of cartilage repair outcomes. UBM provides a three-dimensional image of the cartilage surface, extracellular matrix structure and the subchondral bone with a spatial resolution in the tens-of-microns range. Several experimental studies have demonstrated the feasibility of high-frequency ultrasound to characterise articular cartilage in terms of thickness (Jurvelin *et al.*, 1995; Töyräs *et al.*, 2003), surface integrity (Schöne *et al.*, 2013) and extracellular matrix composition (Gelse *et al.*, 2010).

Jurvelin *et al.* (Jurvelin *et al.*, 1995) showed that ultrasound-based thickness estimation correlates very well with optical or mechanical reference methods. Töyräs *et al.* (Töyräs *et al.*, 2003) later showed that within one anatomical site the error of thickness estimation is below 4 %, if the correct speed of sound is known. The ultrasound surface parameter of "reflection intensity" and the parameter for "roughness" were shown to decrease and increase, respectively, with the onset of

osteoarthritic degeneration (Schöne *et al.,* 2013). Gelse *et al.* (Gelse *et al.*, 2010) showed that fibrotic cartilage in repair tissue exhibits significantly more ultrasound backscatter signal amplitude than healthy articular cartilage. In the present study, a custom-made UBM was used to acquire 3D ultrasound datasets from cartilage repair tissue 1 or 2 years post-operatively. For the first time, such measurements were performed directly after explantation of the stifle. Based on the obtained data, the cartilage layer was reconstructed in 3D and the former focal defect was quantified volumetrically using analytical shape reconstruction algorithms. The UBM parameters derived from this reconstruction were then compared to six clinically relevant standard scoring systems for gross macroscopy, MRI and histopathological grading.

Materials and Methods

Surgical procedure

This experimental study was a central part of a prospective, double blinded, randomised cartilage repair study (TVV number 42/08) on adult female Merino sheep ($n = 32$). The animals were treated in accordance with applicable animal protection laws (Paragraph 8, Section 1) and authorisation by the local legal representative was granted. The chosen bilateral study design was based upon a well-described (Marquass *et al.*, 2011; Zscharnack *et al.*, 2010) chronic full-thickness cartilage defect model (Hepp *et al.*, 2009) that fulfils the requirements for confirmatory animal trials described by the ICRS (Hurtig *et al.*, 2011). In a first operation, bone marrow was harvested at the iliac crest, both knees of anaesthetised sheep underwent open surgery and one cylindrical critical-sized, full-depth, focal cartilage defect was created using a 7 mm diameter biopsy punch on the weight-bearing area of each medial femoral condyle. Care was taken to avoid bleeding resulting from entering the subchondral bone. After 42 d, the right stifle underwent a second arthrotomy and was treated with an autologous mesenchymal stem cell (MSC) seeded collagen-I-scaffold ($n = 16$) or with a cell-free hyper-dense collagen-I-scaffold ($n = 16$, both Arthro Kinetics AG, Heidelberg, Germany). The contralateral left knee was not reopened and these lesions acted as untreated controls ($n = 32$).

Sample acquisition and scoring

The animals were allowed to move freely until they were sacrificed after 1 ($n = 16$) or 2 ($n = 16$) years. First, the legs were explanted and scanned by standard MRI methods (1.5 T, Avanto Syngo MR B17 with Numaris/4, Siemens, Erlangen, Germany, equipped with a micro-imaging unit) to obtain a modified 2D MOCART grade (only for 1 year follow up; Table 2). The applied sequences are listed in Table 3. The total time for MR imaging *per* joint was 23 min. Then the knees were carefully dissected and the former cartilage defects were assessed using ICRS-CRA and Outerbridge (Table 1) gross macroscopic grading systems followed by ultrasonic examination as described below. For the process of opening the joint and completing the ultrasound measurement, less than 1 h *per* sample was required. The exposed cartilage surface

Table 1. Macroscopic classification systems according to Outerbridge (Outerbridge, 1961) and ICRS-CRA (Cartilage Repair Assessment).

Classification system	Item	Score	Description
Outerbridge		0	Normal cartilage
		1	Cartilage with softening and swelling
		2	A partial-thickness defect with fissures on the surface that do not reach subchondral bone or exceed 1.5 cm in diameter
		3	Fissuring to the level of subchondral bone in an area with a diameter more than 1.5 cm
		4	Exposed subchondral bone
ICRS-CRA	Degree of defect repair	4	In level with surrounding cartilage
		3	75 % repair of defect depth
		2	50 % repair of defect depth
		1	25 % repair of defect depth
		0	0 % repair of defect depth
	Integration to border zone	4	Complete integration with surrounding cartilage, Demarcating border < 1 mm
		3	75 % of graft integrated, 25 % with notable border > 1 mm width
		1	50 % of graft integrated with surrounding cartilage, 50 % with a notable border > 1 mm
		0	From no contact to 25 % of graft integrated with surrounding cartilage
	Macrosopic appearance	4	Intact smooth surface
		3	Fibrillated surface
		2	Small, scattered fissures or cracks
		1	Several, small or few but large fissures
		0	Total degeneration of grafted area
	Maximal cumulative score	12	

Table 2. Items of the modified 2D MOCART classification system (Trattnig *et al.*, 2005).

Item	Score	Description
1 – Degree of defect repair and filling of the defect	20	Complete (on a level with adjacent cartilage)
	15	Hypertrophy (over the level of the adjacent cartilage) Incomplete (under the level of the adjacent cartilage; underfilling)
	10	> 50 % of the adjacent cartilage
	5	< 50 % of the adjacent cartilage
	0	Subchondral bone exposed (complete delamination or dislocation and/or loose body)
2 – Integration to border zone	15	Complete (complete integration with adjacent cartilage)
	10	Demarcating border visible (split-like)
	5	Defect visible < 50 % of the length of the repair tissue
	0	Defect visible > 50% of the length of the repair tissue
3 – Surface of the repair tissue	10	Surface intact (lamina splendens intact)
		Surface damaged (fibrillations, fissures and ulcerations)
	5	< 50 % of repair tissue depth
	0	> 50 % of repair tissue depth
4 – Structure of the repair tissue	5	Homogenous
	0	Inhomogenous or cleft formation
5 – Signal intensity of the repair tissue	15	Isointense
(Dual T2-FSE or 3D-GE-FS)	5	Moderately hyperintense
	0	Markedly hyperintense
6 – Subchondral lamina	5	Intact
	0	Not-intact
7 – Subchondral bone	5	Intact
	0	Edema, granulation tissue, cysts, sclerosis)
Maximal cumulative score	75	

Table 3. Applied MR sequences and settings.

Sequence	FOV [mm]	matrix	slice thickness [mm]	TR [ms]	TE [ms]
T1 TIRM cor	150 × 150	512 × 512	3	4810	43
T1 SE sag	150 × 150	512 × 488	3	502	11
PD+T2 SE sag	180 × 180	448 × 448	3	4500	10/70
T2 TRUFI cor 3D	160 × 160	384 × 384	0.4	10	4.85

**Table 4**. Histopathological classification systems according to O'Driscoll (O'Driscoll *et al*., 1998).

Item	Score	Description
1 – Cellular morphology	4	Hyaline articular cartilage
	2	Incomplete differented mesenchyme
	0	Fibrous tissue or bone
2 – Safranin-O staining of the matrix	3	Normal or nearly normal
	2	Moderate
	1	Slight
	0	None
3 – Surface regularity	3	Smooth and intact
	2	Superficial horizontal lamination
	1	Fissures – 25-100 % of the thickness
	0	Severe disruption, including fibrillation
4 – Structural integrity	2	Normal
	1	Slight disruption, including cysts
	0	Severe disintegration
5 – Thickness	2	100 % of normal adjacent cartilage
	1	50-100 % of normal cartilage
	0	0-50 % of normal cartilage
6 – Bonding to adjacent cartilage	2	Bonded at both ends of graft
	1	Bonded at one end, or partially at both ends
	0	Not bonded
7 – Hypocellularity	3	Normal cellularity
	2	Slight hypocellularity
	1	Moderate hypocellularity
	0	Severe hypocellularity
8 – Chondrocyte clustering	2	No clusters
	1	< 25 % of the cells
	0	25-100 % of the cells
9 – Freedom from degenerative changes in adjacent cartilage	3	Normal cellularity, no clusters, normal staining
	2	Normal cellularity, mild clusters, moderate staining
	1	Mild or moderate hypocellularity, slight staining
	0	Severe hypocellularity, poor or no staining
Maximal cumulative score	24	

Table 5. ICRS-I histopathological classification systems (Mainil-Varlet *et al*., 2003).

Item	Score	Description
I – Surface	3	Smooth/continuous
	0	Discontinuities/irregularities
II – Matrix	3	Hyaline
	2	Mixture: hyaline/fibrocartilage
	1	Fibrocartilage
	0	Fibrous tissue
III – Cell distribution	3	Columnar
	2	Mixed/columnar-clusters
	1	Cluster
	0	Individual cells/disorganised
IV – Cell population viability	3	Predominantly viable
	1	Partially viable
	0	< 10 % viable
V – Subchondral bone	3	Normal
	2	Increased remodelling
	1	Bone necrosis/granulation tissue
	0	Detached/fracture/callus at base
VI – Cartilage mineralization (calcified cartilage)	3	Normal
	0	Abnormal/inappropriate location
Maximal cumulative score	18	

Table 6. ICRS-II histopathological classification systems (Mainil-Varlet *et al.*, 2010).

Item	Score	From 0	to	100
1 – Tissue morphology (viewed under polarized light)	0-100	Full-thickness collagen fibres		Normal cartilage birefringence
2 – Matrix straining (metachromasia)	0-100	No staining		Full metachromasia
3 – Cell morphology	0-100	No round/oval cells		Mostly round/oval cells
4 – Chondrocyte clustering (4 or more grouped cells)	0-100	Present		Absent
5 – Surface architecture	0-100	Delamination, or major irregularity		Smooth surface
6 – Basal integration	0-100	No integration		Complete integration
7 – Formation of a tidemark	0-100	No calcification front		Tidemark
8 – Subchondral bone abnormalities/marrow fibrosis	0-100	Abnormal		Normal marrow
9 – Inflammation	0-100	Present		Absent
10 – Abnormal calcification/ossification	0-100	Present		Absent
11 – Vascularization (within the repair tissue)	0-100	Present		Absent
12 – Surface/superficial assessment	0-100	Total loss or complete disruption		Resembles intact articular cartilage
13 – Mid/deep zone assessment	0-100	Fibrous tissue		Normal hyaline cartilage
14 – overall assessment	0-100	Bad (fibrous tissue)		Good (hyaline cartilage)
Cumulative score (mean)	0-100			
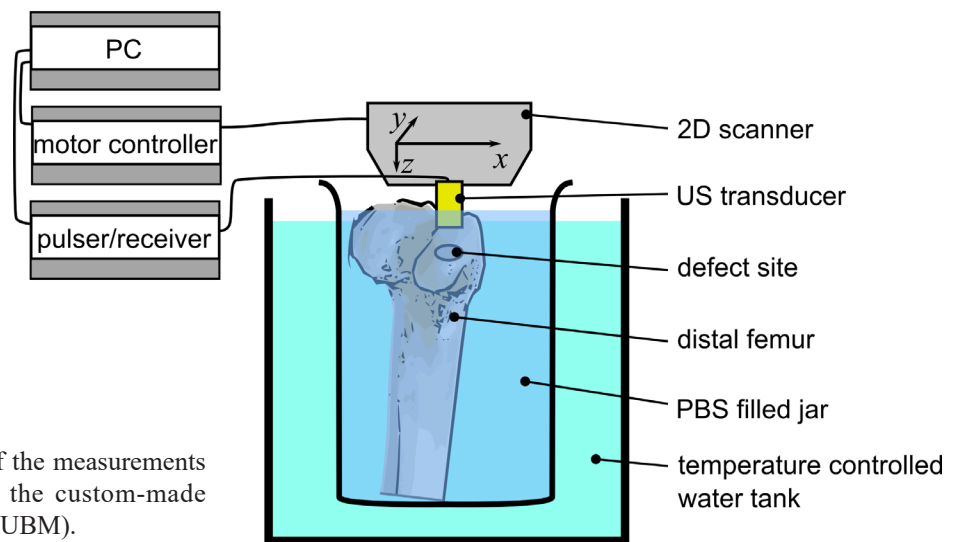
Fig. 1. Schematic set-up of the measurements on a femur condyle with the custom-made ultrasound biomicroscope (UBM).

was moistened with phosphate buffered saline (PBS, Biochrom, Berlin, Germany) at all times. Subsequently, the samples were fixed in 4 % formalin (Roth, Karlsruhe, Germany). For histological preparation the condyles were cut centrally through the repair site into two halves such that the section included the worst apparent outcome. Two histological slides of a maximum thickness of 45 µm were then prepared from one half block in routine fashion by the cutting-grinding technique according to a modified Technovit 9100 protocol (Willbold and Witte, 2010) based on the methyl methacrylate-based embedding resin system. Histopathological grading according to O'Driscoll (Table 4), ICRS-I (Table 5), and ICRS-II (Table 6) scoring systems was performed by two experienced pathologists. All examinations were done in blinded fashion by the radiologist, surgeon, medical physicist and both pathologists.

Ultrasound biomicroscopy
A custom portable ultrasound biomicroscope (UBM) was developed by an experienced ultrasound group (Gelse *et al.*, 2010; Schöne *et al.*, 2013) for the sonographic evaluation (Fig. 1). It consists of an industrial computer (ACME II, ACME Portable Machines, Azusa, CA, USA) equipped with an AD-card (CompuScope 14200, Gage Applied Technologies Inc., Lachine, QC, Canada) and a 2-axis-controller card, an external pulser-receiver (DPR500, JSR Ultrasonics, Pittsford, NY, USA) and an external 2D-scanner unit mounted with a spherically-focused, single-element lithium niobate transducer (f-number: 2.66, aperture diameter: 3 mm; NIH Resource Center for Ultrasonic Transducer Technology, Los Angeles, CA, USA). Custom made software (MATLAB 2009b, The Mathworks Inc., Natick, MA, USA) was used. The 2D scanner provided a scan increment of 22.4 µm and 24.9 µm in x- and y-directions, respectively, and a total scan area of 12.5 × 9.5 mm. The transducer provided resolutions of 50 µm and 120 µm in axial and lateral directions, respectively. The centre frequency was 37.5 MHz and the −6 dB range was 25-50 MHz, which corresponds to a fractional −6 dB bandwidth of 65 %. Focus distance and focal length were 9 mm and 3.3 mm, respectively. The spatial and spectral transducer characteristics were measured using the wire method (Raum and Obrien, 1997).

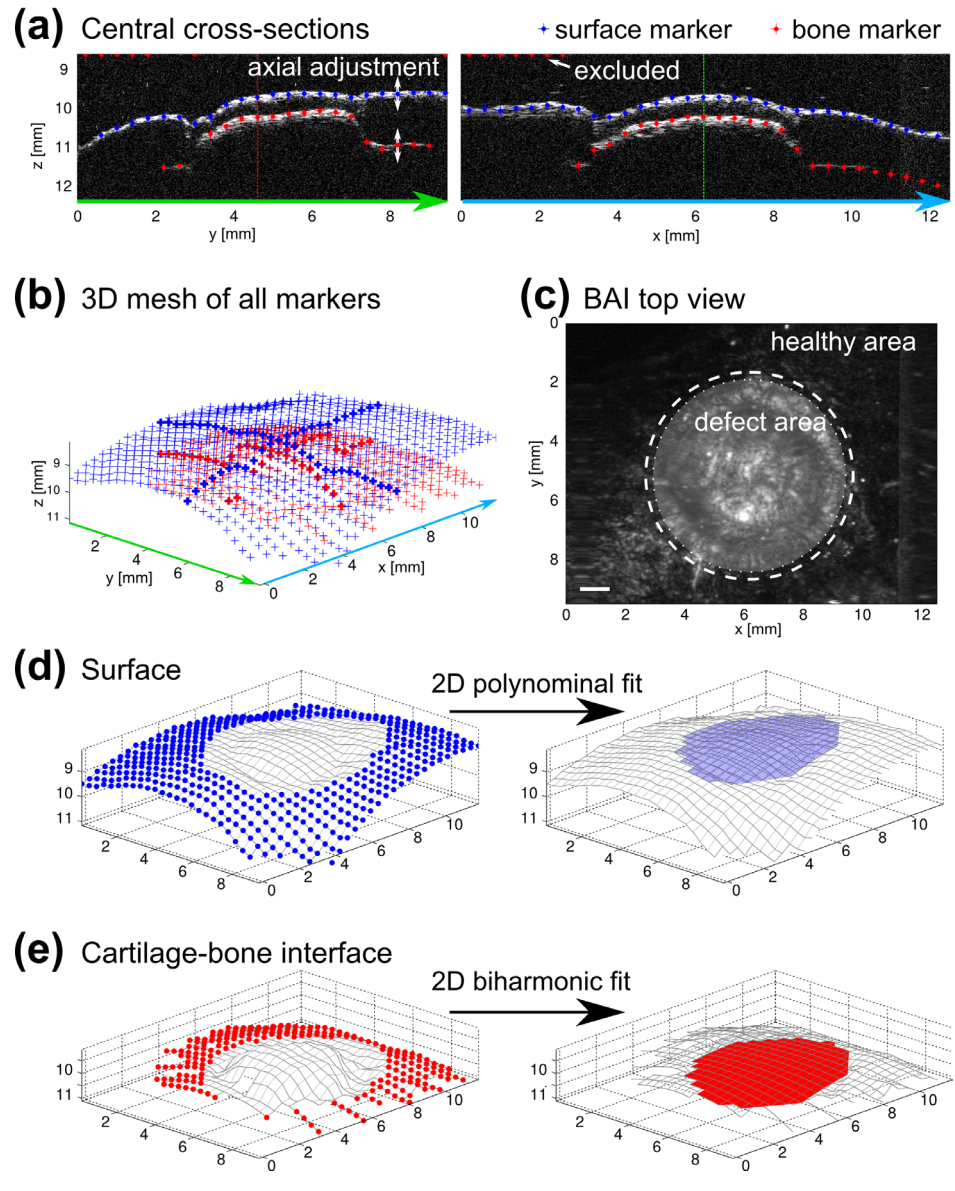For the measurements the femoral condyles were completely immersed in a tank filled with phosphate

Fig. 2. Example of interface fitting of cartilage surface (blue) and cartilage-bone interface (red). (**a**) Two central orthogonal cross-sections with the markers for cartilage surface and cartilage-bone interface. (**b**) 3D visualisation of all surface and cartilage-bone interface markers. The bold markers correspond to the cross-sections of Fig. 2**a**. (**c**) Ultrasound BAI top view image and the manual positioned circle of 7 mm diameter to separate the healthy and defect area. (**d-e**) Markers of the healthy area (left) are used to estimate the original surface (**d**) and cartilage-bone interface (**e**) in the defect area (right).

buffered saline (PBS, Biochrom, Berlin, Germany) at 36.1 ± 0.9 °C, and aligned such that the former defect area was oriented upwards. The scanner unit was positioned above the sample and oriented such that the scan area included the defect region and the surface pulse-echo appeared between 11.0 and 11.5 µs, representing the begin of the focus. The correct alignment was checked with live cross sectional B-mode scans. Finally, a C-scan (*i.e.* a two-dimensional scan in a scan plane parallel to the defect surface) was performed and the digitised pulse-echoes were stored for later offline processing.

The acquired pulse-echoes were evaluated with custom-made software in MATLAB (The Mathworks Inc., Natick, MA, USA). An overview image was computed by integration of each envelope curve for every pulse-echo (Backscattered Amplitude Integral – BAI (Raum *et al.,* 1998)). The surface position was detected in two steps as described previously (Schöne *et al.,* 2013). Briefly, for each pulse-echo a threshold-based detection was performed on the envelope curve obtained by Hilbert transform, providing a simple approximation of the surface

reflection time-of-flight. In the second step the surface time-of-flight map was smoothed by a custom LOWESS filter (locally weighted scatter plot smoothing) (Cleveland, 1979), modified for usage on two-dimensional maps. The conversion from time domain to spatial domain was performed by usage of the PBS sound velocity of 1540 m/s at 36 °C. The settings for threshold and filter span were verified manually at 2D cross-sections and varied between −52 dB to −47 dB and 3 % to 6 %, respectively.

The cartilage-bone interface was detected by locating the maximum below the surface peak of the envelope curve averaged over 8 adjacent (in x-direction) pulse-echoes. A fixed cartilage speed of sound of 1660 m/s was assumed for all samples for the conversion from time of flight into distance. Both, the surface map and the cartilage-bone interface map were resampled on a coarser grid with a grid size of 400 µm in the x-y-plane, to reduce the total number of points *per* grid to 744 (Fig. 2**a-b**).

In manual and blinded post processing, the automatically detected surface and cartilage-bone interface positions of the coarse grid were controlled manually by visual
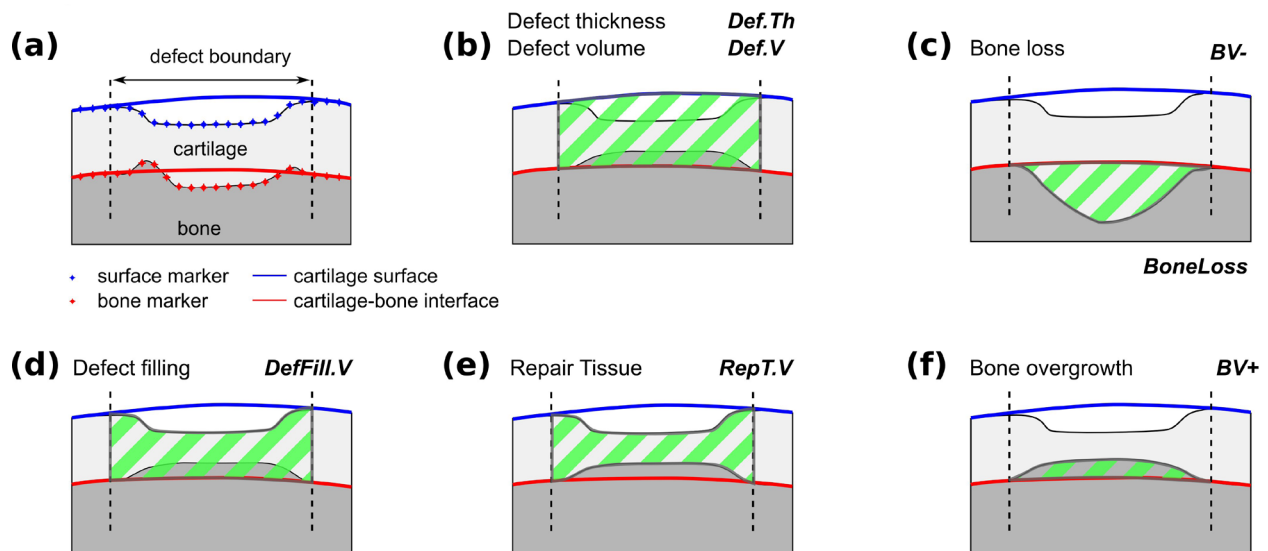
Fig. 3. (**a**) Schematic cross sections of cartilage defects as seen in the ultrasound images including the markers and the estimated interfaces. (**b**-**f**) Schematic explanation of the calculation of the derived parameters.

inspection of successive cross-sectional images and adjusted if necessary (Fig. 2**a**). This was done by a clinician, subsequently checked by an ultrasound expert. Furthermore, the defect region was estimated by adjusting a 7 mm diameter circle to fit best to the ultrasound BAI top view image (Fig. 2**c**). For the following calculations the inner part of the circle excluding a fringe of 0.25 mm was defined as "defect area" and the outer part as "healthy area" (Fig. 2**c**).

Interface fitting
Femoral condyles in general have a convex shape. To estimate the original cartilage surface in the defect area, the surface markers of the surrounding healthy area were used to apply a 2D 4th order polynomial fit excluding the 3rd order terms. The average number of surface markers for fitting was 372 ± 34 (Fig. 2**d**).

The ultrasound signal of the cartilage-bone interface was often much weaker than that of the cartilage surface. Therefore, a more robust fitting was required. The original cartilage-bone interface was assumed by a 2D biharmonic interpolation (built-in function of MATLAB) of the bone positions of the healthy area. The average amount of bone markers for fitting was 254 ± 69 (Fig. 2**e**).

Calculation of thickness and volumetric parameters
All calculations for thickness and volumetric parameters were based on the coarse grid which comprised on average 227 ± 28 points in the defect area. Based on the surface and bone markers and the estimated shapes of the original cartilage surface and bone layer (Fig. 3**a**), the defect thickness *Def.Th* and defect volume *Def.V* was calculated (Fig. 3**b**). For each point inside the defect area, the thickness values were estimated from the difference between the estimated cartilage surface to the estimated cartilage-bone interface. Based on this, the minimum and maximum values were extracted and the average thickness was calculated.

The subchondral bone plate is often affected during cartilage repair and loses its initial shape. Either the subchondral bone grows above its initial level, or bone is lost, *e.g.* caused by cyst formation. In accordance with Gomoll *et al.* (Gomoll *et al.*, 2010), we defined the volume between actual bone and initial cartilage-bone interface as bone loss (*BV*)− (Fig. 3**c**). All tissue above the initial cartilage-bone interface and the actual surface was denoted as defect filling, *DefFill.V* (Fig. 3**d**). This was further differentiated into the soft repair tissue, *RepT.V* (Fig. 3**e**) and bone overgrowth, *BV+* (Fig. 3**f**). The parameter abbreviations are based on the system of Parfitt *et al.* (Parfitt *et al.*, 1987) for bone histomorphometry. For comparison between the samples the parameters *DefFill.V*, *RepT.V* and *BV+* were normalised to the initial defect volume *Def.V*.

Statistical Analysis
Statistical analysis was performed in cooperation with a professional statistician (ACOMED statisik, Leipzig, Germany). Kruskal-Wallis tests followed by Mann-Whitney U *post-hoc* analysis were performed to identify differences between single items of a classification system. In the provided boxplots horizontal lines between groups indicate significant differences. Additionally, for all classification systems, except for Outerbridge, the cumulative scores were calculated by adding up the single values of each item. For ICRS-II the average value of all 14 items was taken. Correlations between the volumetric UBM data and the cumulative scores were expressed according to Spearman's rank correlation. In accordance with Blackman *et al.* (Blackman *et al.*, 2013), good to excellent correlation was defined as a ρ value > 0.75, moderate to good 0.5-0.75, weak 0.25-0.5, and little or no correlation < 0.25. Bonferroni-Holm correction was used with the global significance level of 0.05.

In the provided scatter plots with linear correlations dashed and dotted lines correspond to the 95 % confidence interval and 95 % prediction interval, respectively.
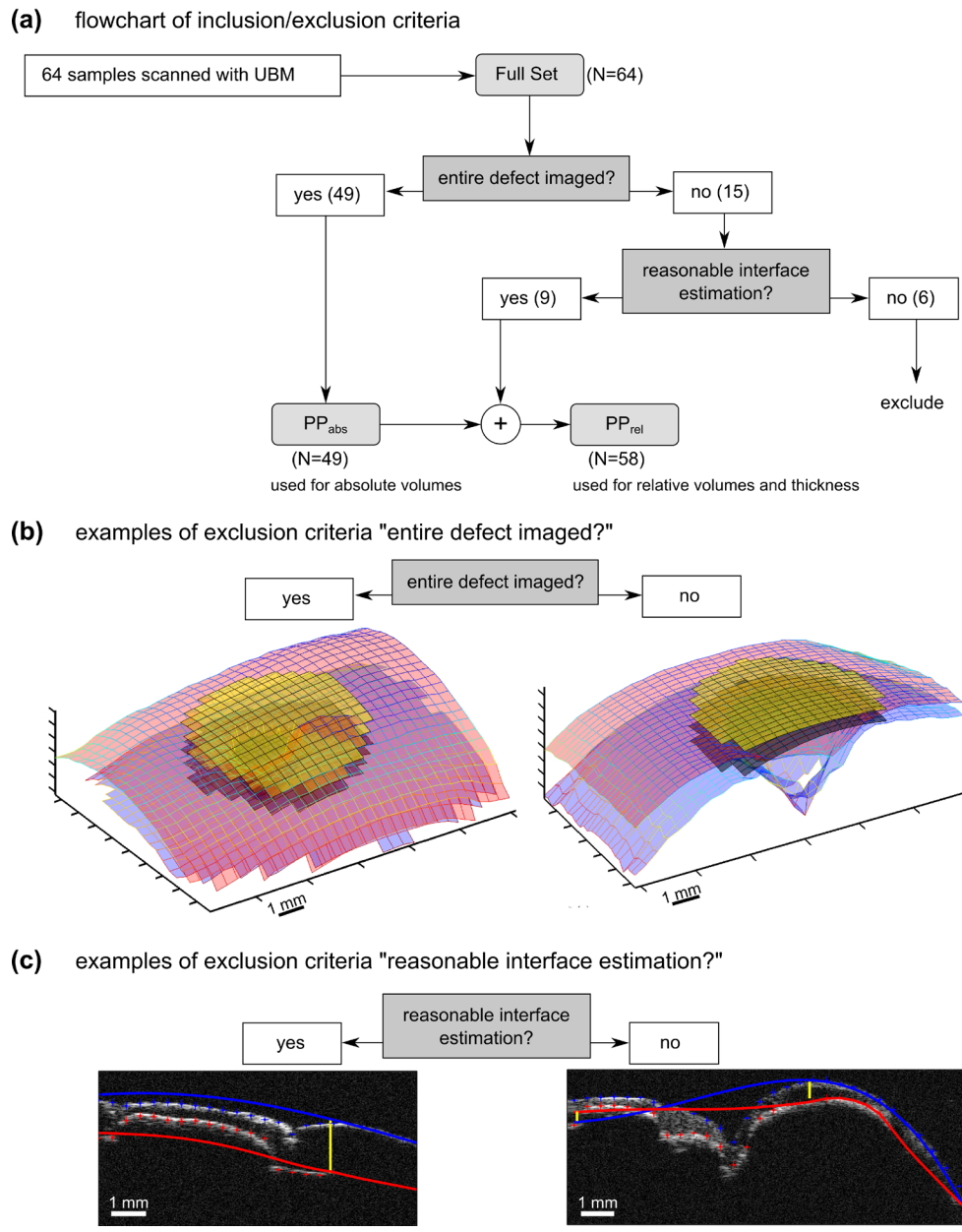
(a)   flowchart of inclusion/exclusion criteria

(b)   examples of exclusion criteria "entire defect imaged?"

Fig. 4. (**a**) Flowchart of inclusion/exclusion criteria. (**b-c**) Four different samples to demonstrate the inclusion/exclusion criteria. (**b**) 3D reconstructions of the interfaces of cartilage and bone as well as the estimated interfaces at the defect region. The right figure clearly shows that a peripheral part of the former cartilage defect was not acquired. (**c**) 2D cross-sections of the central region with reasonable (left) and not reasonable (right) interface fitting.

(c)   examples of exclusion criteria "reasonable interface estimation?"

Inclusion/exclusion criteria

The inclusion/exclusion scheme is shown in Fig. 4**a**. The first criterion was "entire defect imaged?" and, if not, the second one was "reasonable interface fitting?". Fulfilment of the first criterion allowed reliable estimation of all parameters; these samples were denoted as PP_{abs}. Failure of the first, but fulfilment of the second criterion allowed calculation of the thickness values and the relative volumetric parameters, but not the absolute volumes. These samples are denoted as PP_{rel}. Failure on both criteria led to exclusion of any evaluation ($n = 6$). All provided values of cartilage thickness (mean, min, max) and the relative volumes of *DefFill.V*, *RepT.V*, and *BV+* were calculated from $PP_{abs} + PP_{rel}$ ($n = 58$), while the absolute values of *Def.V* and *BV−* were calculated from PP_{abs} ($n = 49$). MR images were only available for comparison for the samples of 1-year follow-up and with the applied inclusion/exclusion criteria, 26 samples remained for comparison of

volumetric data to the modified 2D MOCART subscores. Examples for fulfilment or failure of both criteria are shown in Fig. 4**b-c**.

Results

Comparison of histopathological to ultrasound cross-sections

The 3D dataset of the UBM scans allowed for reconstruction of cross-sectional images in arbitrary directions. By matching the structures of the cartilage surface and subchondral bone we reconstructed an ultrasound cross-sectional image for each sample showing the same section as the corresponding histological slides.

Fig. 5**a** shows a sample with a cyst in the central former defect region surrounded by bone overgrowth. Of note, the final histological slides do not include the entire cyst.
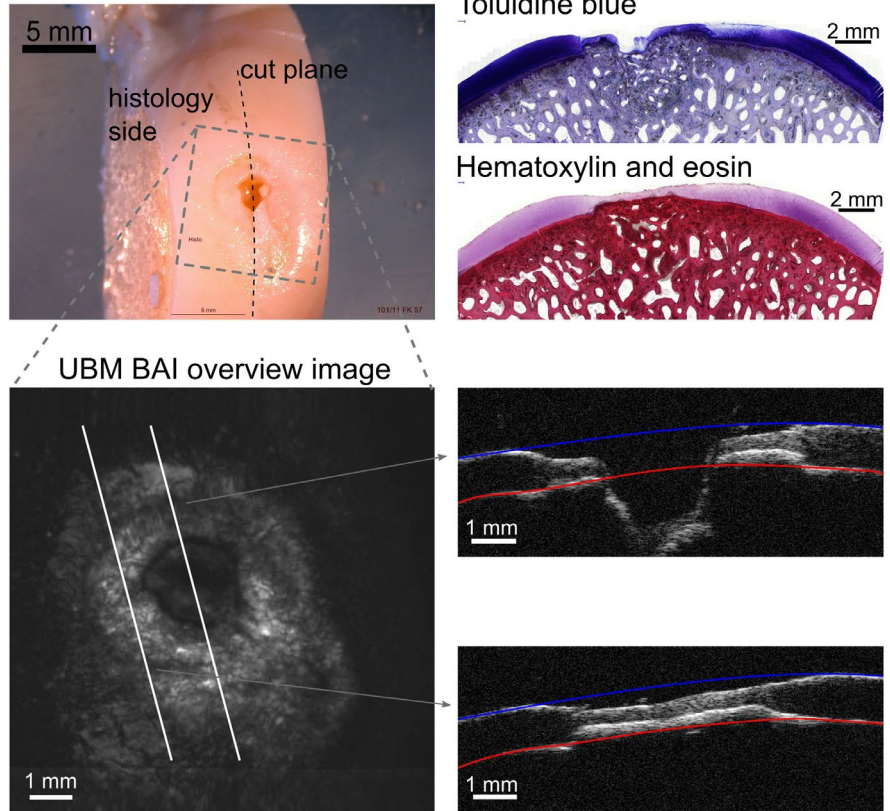
Fig. 5. Comparison of histological and ultrasound cross-sectional images. (**a**) Macroscopic view on sample after cutting centrally through the defect region. The left part was used for histology, but both slides do not include the central cyst. Ultrasound cross-sectional images are just 1.5 mm apart. (**b**) Sample with deep cyst where the deeper cyst structures cannot be visualised by ultrasound. The open bone lamella indicates deep cyst. (**c**) Huge cyst covered by subchondral lamella is not visible in ultrasound; the breach can only be seen in the appropriate slide. (**d**) Sample of exposed bone and superficial damage of subchondral lamella. The dashed and solid vertical lines indicate the 7 mm diameter defect boundary and the additional 0.25 mm fringe, respectively. (**e**) Based on the derived scores, this was the best outcome of the present study. However, the sample still contains a large amount of bone overgrowth and a cyst.
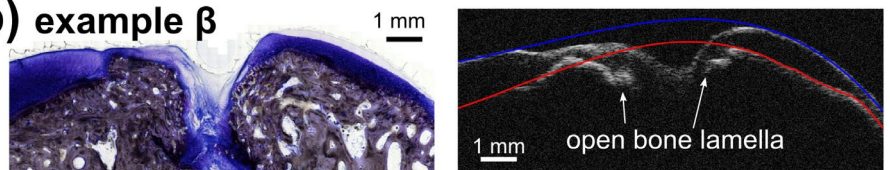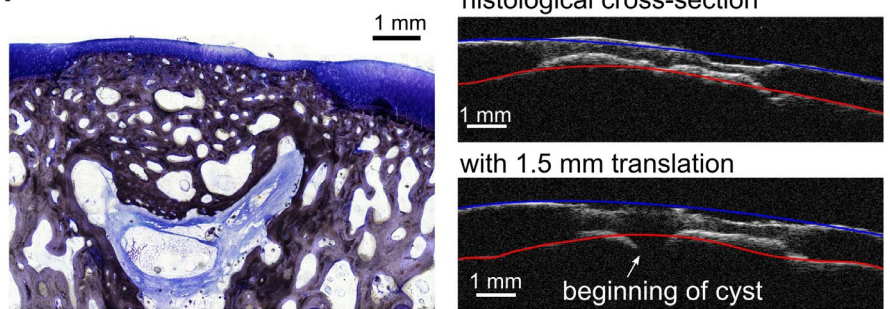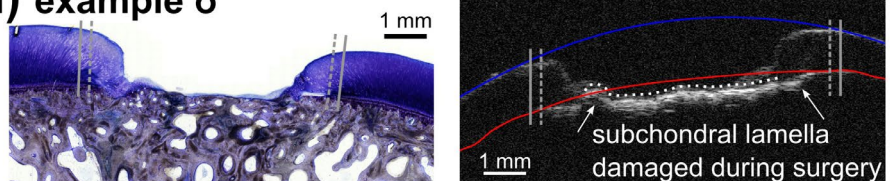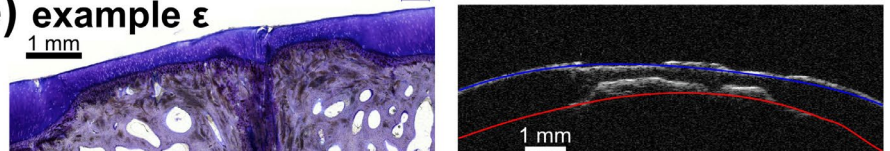
Table 7. Descriptive statistics of estimated thickness and volumetric parameters derived from UBM for all samples.

Value		unit	mean		std
Thickness *Def.Th*	minimum	[mm]	0.67	±	0.25
	mean	[mm]	1.01	±	0.25
	maximum	[mm]	1.36	±	0.22
Def.Th difference (left – right)	mean	[mm]	0.07	±	0.26
	absolute value	[mm]	0.20	±	0.17
Volumetric parameters	*Def.V*	[mm³]	32.85	±	7.20
	DefFill.V	[%]	71.92	±	16.58
	RepT.V	[%]	62.92	±	12.46
	BV+	[%]	9.00	±	7.83
	BV-	[mm³]	5.22	±	6.02

Table 8. Cross tabulation of macroscopic, MR and histopathological scores and UBM parameters for defect filling *DefFill.V* and repair tissue *RepT.V*. Spearmans ρ is provided in **bold**, p values in normal font style.

		n	ICRS-CRA	Modified 2D MOCART	O'Driscoll	ICRS-I	ICRS-II	ICRS-II "overall assessment"	UBM DefFill.V	UBM RepT.V
Outerbridge		58	**-0.679**	**-0.738**	**-0.715**	**-0.725**	**-0.723**	**-0.718**	**-0.633**	**-0.386**
			<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0,003
ICRS-CRA	cumulative score	58		**0.475**	**0.549**	**0.534**	**0.691**	**0.636**	**0.582**	**0.361**
				0,014	<0.001	<0.001	<0.001	<0.001	<0.001	0,005
Mod. 2D MOCART	cumulative score	26			**0.731**	**0.816**	**0.603**	**0.644**	**0.720**	**0.560**
					<0.001	<0.001	0,001	<0.001	<0.001	0,003
O'Driscoll	cumulative score	58				**0.912**	**0.824**	**0.847**	**0.548**	**0.315**
						<0.001	<0.001	<0.001	<0.001	0,016
ICRS-I	cumulative score	58					**0.844**	**0.836**	**0.595**	**0.406**
							<0.001	<0.001	<0.001	0,002
ICRS-II	cumulative score	58						**0.940**	**0.715**	**0.486**
								<0.001	<0.001	<0.001
ICRS-II	"overall assessment"	58							**0.647**	**0.373**
									<0.001	0,004
UBM	*DefFill.V*	58								**0.846**
										<0.001
UBM	*RepT.V*	58								

Only the toluidine blue stained slide includes a small gap representing a marginal part of the cyst. Two ultrasound cross-sections with 1.5 mm distance were reconstructed to demonstrate the remarkable differences that can occur with such a small shift. One central cross-section shows the full extension of the cyst central to the defect involving the bone approximately 2 mm deep. The peripheral cross-section shows repair tissue filling almost the entire defect and subchondral overgrowth – very similar to the haematoxylin and eosin (HE) slide. Of note, the repair tissue shows high ultrasound backscatter, whereas the surrounding hyaline cartilage is anechoic.

A very deep cyst is shown in Fig. 5**b**. The ultrasound cross-section cannot visualise the full depth of the cyst. However, the not closed subchondral lamella indicates that the hole progresses into the bone.

In Fig. 5**c** a sample with a deep cyst covered by the subchondral bone is shown. The cyst tissue cannot be visualised with ultrasound and therefore its volume cannot

be estimated reliably. The breach of the cyst through the subchondral bone can be found in the appropriate ultrasound cross-sections.

Fig. 5**d** shows a sample with an empty defect and exposed subchondral bone. Looking closer, slight damage to the subchondral lamella can be seen, probably caused by the initial surgery. In the ultrasound cross-section, the empty defect shows increased signal intensity at the exposed bone because no attenuating tissue is covering it. Furthermore, the bone layer is located below the estimated initial bone interface, caused by the damage of the first surgery. It should be noted, this shift in the ultrasound image is further enhanced by the speed of sound differences between cartilage tissue and the PBS filled empty defect, whereas for the visualisation a constant speed of sound is assumed. This assumption does not affect the volumetric parameters, since they were calculated with the material-specific sound velocities. In the ultrasound cross-section image of Fig. 5**d** the corrected position of the bone layer is
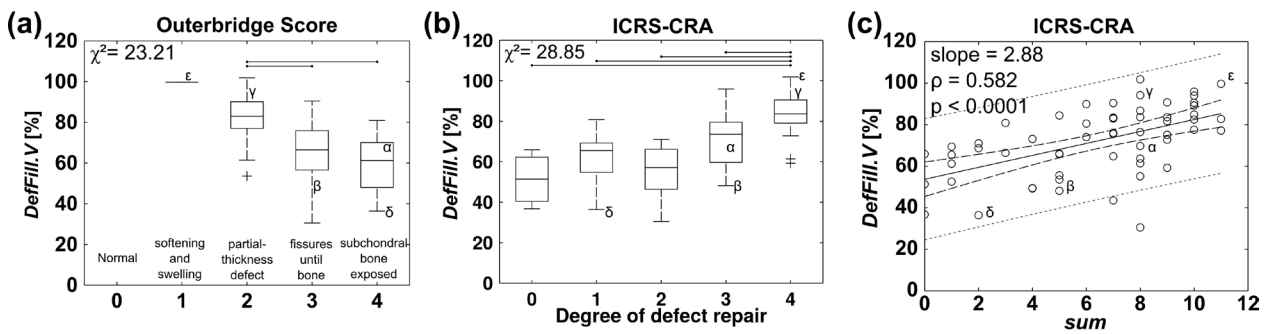
Fig. 6. Correlations of macroscopic scores to the defect filling *DefFill.V*. Boxplots of *DefFill.V* grouped according to (**a**) the Outerbridge Score and (**b**) the ICRS-CRA item "Degree of defect repair". (**c**) Linear correlation between *DefFill.V* and the cumulative score of ICRS-CRA.
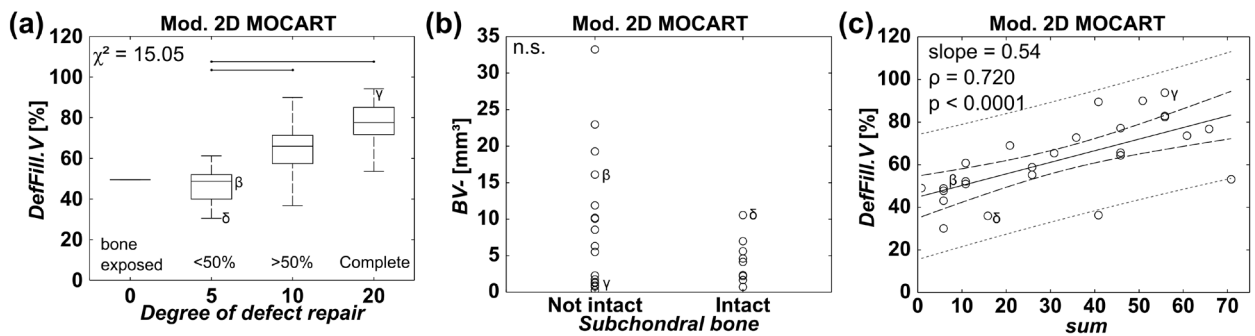

Fig. 7. Correlations of modified 2D MOCART items to *DefFill.V* and *BV−*. (**a**) Boxplots for *DefFill.V* grouped according to the item "Degree of repair tissue". (**b**) The volume of bone loss *BV−* grouped as "Not intact" and "Intact" according to the item "Subchondral bone". (**c**) Linear correlation between *DefFill.V* and the cumulative score of modified 2D MOCART.

marked by the dotted line. This was calculated manually for this specific sample to visually demonstrate the dimension of the effect. Additionally, this example shows that the cartilage adjacent to the defect is bent into the defect region when the lateral mechanical support is lost. The tissue bent into the defect region is counted as tissue filling the defect in the ultrasound data evaluation.

The sample with the best outcome of the present study is shown in Fig. 5**e**. The repair tissue is filled up to the level of the surrounding tissue. However, a large amount of the initial cartilage height is filled by bone overgrowth. The obtained ultrasound parameters are: *DefFill.V*: 99.7 %; *RepT.V*: 75.4 %; *BV+*: 24.3 %. The related scores are the following: Outerbridge: 1; ICRS-CRA "degree of defect repair": 4; ICRS-CRA sum: 11 (max. 12); O'Driscoll "thickness": 0; O'Driscoll sum: 20 (max. 24); ICRS-I informative sum: 18 (max. 18); ICRS-II mean: 83 (max. 100); ICRS-II "overall assessment": 60 (max. 100).

In Figs. 6-8 the position of the examples, shown in Figs. 5 **a**, **b**, **c**, **d** and **e**, are marked by Greek letters α, β, γ, δ and ε.

Statistical analysis
The results of cartilage minimum, mean, and maximum thickness, thickness differences between left and right knees, defect volume and the derived volumetric parameters are listed in Table 7. The intra-rater correlation and the intra-class correlation (ICC) for O'Driscoll, ICRS-I, and ICRS-II were $r \geq 0.93$ ($p < 0.0001$) and

$ICC \geq 0.95$, respectively. The correlation parameters between Outerbridge, the cumulative scores of ICRS-CRA, modified 2D MOCART, O'Driscoll, ICRS-I, ICRS-II, the single item ICRS-II "overall assessment", and the UBM parameters *DefFill.V* and *RepT.V* are shown in Table 8.

As the present study was intended to compare UBM parameters to established classification systems, no results regarding the surgical treatment are shown.

The relation between *DefFill.V* and the Outerbridge scores and the ICRS-CRA item "Degree of defect repair" are shown in Figs. 6**a** and 6**b**, respectively. Fig. 6**c** shows the correlation between the cumulative ICRS-CRA score and *DefFill.V*.

The modified 2D MOCART item 1 "Degree of defect repair" and its comparison to *DefFill.V* is shown in Fig. 7**a**. In total, 9 samples were assessed as completely filled (20 points). The 2D MOCART item "Subchondral bone" can either be "intact" or "not intact – oedema, granulation tissue, cysts, sclerosis" and therefore was expected to show differences in *BV+* or *BV−* volumes. However, no significant differences were found. When looking at the raw data (Fig. 7**b**), the five samples with the highest *BV−* values were grouped as "not intact". Notably, the "not intact" group also includes several samples with low *BV−* values. Fig. 7**c** shows the correlation between the cumulative modified 2D MOCART score and the *DefFill.V*.

The comparisons between *DefFill.V* and O'Driscoll item 5 "Thickness" and the cumulative O'Driscoll score are shown in Figs. 8**a** and 8**b**, respectively.
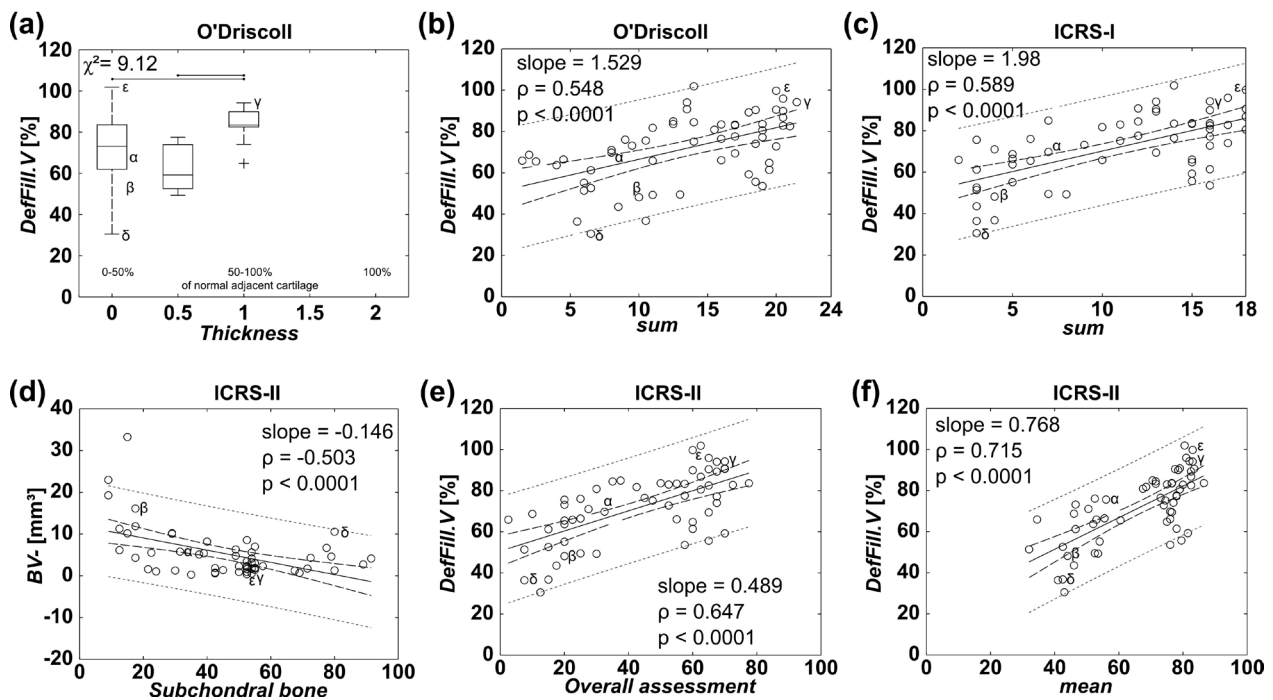
Fig. 8. Correlations of histopathological classification systems and defect filling *DefFill.V* or bone loss *BV−*. (**a**) Boxplot of *DefFill.V* grouped according to the O'Driscoll item "Thickness". (**b-c**) Linear correlation between *DefFill.V* and the cumulative scores of (**b**) O'Driscoll and (**c**) ICRS-I. (**d**) Negative linear correlation between the *BV−* and the ICRS-II item "subchondral bone". (**e-f**) Linear correlation between *DefFill.V* and (**e**) the ICRS-II item "overall assessment" and (**f**) the mean of all items of ICRS-II.

Both histopathological ICRS classification systems have no item addressing defect filling explicitly. Item 5 of ICRS-I "Subchondral bone" was found to have a weak, significant correlation to *BV+* ($\rho = 0.286$, $p < 0.03$).

The correlation between *DefFill.V* and the cumulative ICRS-I score is shown in Fig. 8**c**. ICRS-II item 8 "Subchondral bone abnormalities/marrow fibrosis" showed no correlation with the *BV+*, but moderate, significant negative correlation with the *BV−* was noted (Fig. 8**d**). The correlations between *DefFill.V,* the ICRS-II item "overall assessment" and the cumulative ICRS-II score are shown in Figs. 8**e** and 8**f**, respectively.

Discussion

In this study the feasibility of UBM as an objective method for the evaluation of cartilage repair outcomes was investigated on a large animal model.

Post-processing of the UBM data was performed semi-automatically. User input was needed at two steps of the post-processing: i) to define the centre of the defect area and ii) to correct for surface and bone markers which were occasionally set incorrectly by the software. The surface detection generally worked well. More problems occurred with detection of the cartilage-bone boundary. Manual re-positioning was usually required at sites of irregular ossification, at cyst tissue, and at outer regions where the signal was too low caused by the large tilt. However, the volumetric quantification based on UBM can be seen as an objective method and therefore only systematic errors could have an influence on the result.

One of these potential error sources is the unknown exact speed of sound influencing the positioning on the z-axis. As the speed of sound in the coupling medium is usually well known, the z-position of the cartilage surface can be assumed as very precise. More uncertainty exists for the speed of sound in cartilage, and especially for cartilage repair tissue, affecting the z-positioning of the cartilage-bone interface. This directly influences the derived volumetric parameters. As previously shown, at a constant site the variations of speed of sound in healthy articular cartilage are rather small; *e.g.,* 2 % for the medial femoral condyle and trochlea, 4 % at the patella (Töyräs *et al.*, 2003). To the knowledge of the authors, speed of sound for cartilage repair tissue has not previously been documented. However, it is known that repair tissue is usually less stiff (Gelse *et al.*, 2010), therefore the speed of sound is probably reduced like in degenerated cartilage (Töyräs *et al.*, 2003). The lowest value for degenerated cartilage found by Töyräs *et al.* (Töyräs *et al.*, 2003) was 1560 m/s. Since we used a constant value of 1660 m/s, the maximal expected error for repair tissue speed of sound is 7 %. However, due to the use of a constant speed of sound it is a systematic error that is equally distributed over all samples and probably does not affect the correlations to the established classification systems.

A second systematic error could occur due to flattening of the healthy cartilage surrounding the defect. Due to the increased mechanical stress after the defect was created, the cartilage thickness may decrease around the defect. Since the shape of this tissue was used to estimate the initial cartilage surface, a flattening would cause a decrease of the estimated thickness and defect volume. To overcome

potential influences of flattened surrounding cartilage, measurements should include sufficient healthy tissue around the defect area. In the present study a 0.25 mm fringe around the defect area was used. When looking at Fig. 5**d**, this seems to be enough for the defects of this study, at least to have no influence of bent tissue. However, it is hard to judge possible flattening based on the present ultrasound data, as the field of view is likely too small for that. The lateral extension of flattening can easily be judged by the naked eye based on circular light reflections from the glossy cartilage around the defect (Jackson *et al.*, 2001). Large scale histology slides could allow assessment of potential flattening, at least in two dimensions.

A wide range of repair outcomes were observed, from deep cysts over exposed subchondral bone to defects filled to the level of the surrounding cartilage. A high variation of cartilage thickness was found between all samples, but also between the left and right knees of the individual sheep (Table 7). This is caused by natural variation but also by little, unintentional variations of the anatomical location used for the defect. On the one hand, the finding that the intra-sheep variation (0.20 mm) is almost as substantial as the inter-sheep variation (0.25 mm) supports the use of the bilateral model (Hurtig *et al.*, 2011). On the other hand, the high variations in cartilage thickness cause problems when creating the defect. Since the cartilage thickness is usually not measured before and a constant defect depth is used, the resulting defect could be ICRS Grade 2, 3 A, B, C, 4 A or B, depending on the depth, whether the calcified layer is intact or not and if the subchondral bone is damaged. The repair potential and outcomes of partial or full-thickness cartilage defects and defects with damaged subchondral bone plate are entirely different. Therefore, the comparability of such defects in an animal model should be discussed critically. Measuring of cartilage thickness – for example with ultrasound – before creating the individual defect and controlled removal of tissue in depth is highly recommended and would improve further non-clinical studies in terms of consistency.

Also in the present study we could not avoid occasional damage to the subchondral bone plate. However, this was more beneficial than adverse, as this resulted in larger variations between the repair outcomes.

This was reflected in the values of the established classification systems covering the range from low to high scores. Between all six established scores good to excellent significant correlations were found. This supports the validity of our data. Although, ICRS I and II were developed for histology slides prepared from 2 mm biopsies, but applied to histology slides of the whole joint, they showed excellent correlations with the O'Driscoll score.

Where a causal relationship was present, single items of the established scores were directly compared to the UBM parameters, *e.g.* ICRS-CRA "Degree of defect repair" to defect filling *DefFill.V*.

The defect filling assessed by macroscopic evaluation is graded according to the Outerbridge score and ICRS-CRA item 1 "Degree of defect repair". The Outerbridge score was the only score in this study where the lowest value equals the best outcome; therefore, it shows a negative correlation with *DefFill.V*. Except for one sample (ε), all samples were scored with Outerbridge score 2-4 (partial thickness defect – exposed bone). The common problem for macroscopic evaluation of cartilage repair is the impossibility of assessing structures below the surface. Total or almost total filling of the defects can easily be assessed (examples γ and ε). However, partial filling can hardly be expressed as a relative value as long as the initial cartilage thickness is unknown. Therefore it is not surprising that no differences for *DefFill.V* were found for ICRA-CRA "Degree of defect repair" scores 0-3. Depending on the definition, the categorical grading can cause overestimation of some features – such as grade 4 of the Outerbridge score. Example α has partially exposed subchondral bone, but in total a defect filling of 60 %. However, it is in the same category like a total empty defect (δ). In contrast, the 3D visualisation with UBM can also show the morphology below the surface. Therefore, the absolute thickness of the repair tissue can be assessed and compared to adjacent hyaline cartilage. Furthermore, malformations of the subchondral bone plate can be visualised in detail.

The modified 2D MOCART item "Degree of defect repair" seems to be positively biased, since 35 % of the samples were assigned to have "complete" filling. In *DefFill.V* these samples reached values from 50 % to 95 %. Unfavourable is the definition of the categories "filling above 50 %" and "complete filling". This causes observer bias and, in combination with insufficient resolution, an information bias. Clinical MRI is the most sophisticated *in vivo* imaging modality, but it has limited capability for a quantitative volumetric evaluation of the repair tissue due the limited spatial resolution and slice thicknesses of 2-3 mm. The resolution appears to be too poor for the small structures we were examining during this study. Since human cartilage is around three times thicker than ovine cartilage, more details could be visualised with the same resolution. Thereby, an even better correlation to the defect filling *DefFill.V* is expected. However, at this point we cannot extrapolate the MRI results of the present study to human application. To do so, direct comparison of UBM and MRI on human would be required. Experimental high-field MRI at 9.4 T has been shown to allow reliable evaluation of explanted cartilage repair tissue in sheep, since it provides much better resolution (Goebel *et al.*, 2014).

In histopathological grading the O'Driscoll item "thickness" focuses only on the thickness of the cartilaginous layer, but not on the total filling of the defect. As the amount of bone overgrowth is neglected in this item, one of the best (ε) and one of the worst (δ) outcomes were both grouped in the lowest score. No sample of the present study was scored with "100 %" and two-thirds of all samples were assigned to have a thickness below 50 %. In contrast, *DefFill.V* of these samples showed values from 30 % to 100 %, with a median of 73 %.

ICRS-I and ICRS-II are designed for biopsy samples obtained arthroscopically, thus defect filling cannot be assessed. The majority of the items of ICRS-I and ICRS-II had no causal relation to the volumetric parameters obtained from UBM; therefore, no statistical tests were

performed on them. Only the item "Subchondral bone" of ICRS-II correlated negatively with the bone loss $BV-$. The samples with the highest $BV-$ value usually had a score below 20.

Discussion is required about the justification of using ICRS-I and ICRS-II for the evaluation of histology slides showing the entire cartilage repair site and adjacent hyaline cartilage. Basically, all parameters of both classification systems can be assessed based on these slides. However, it is possible, that the score is affected by the more comprehensive impression the rater gets from the repair tissue. Also, different qualities of repair tissue within one sample could cause uncertainty about which part to evaluate, which increases the influence of the rater. However, in the present data we found very good intra-rater and intra-class correlations and excellent correlations to the O'Driscoll score. Therefore, we consider both ICRS histopathological classification systems to be usable in the context of this study. However, caution is advised when comparing the present scores with other studies in which only biopsy samples were evaluated.

The content of histological slides strongly depend on the position where the samples were cut. Since only two individual 2D cross sections of the often complex three-dimensional morphology of a cartilage repair site were available, scoring is highly prone to over- or underestimation of certain aspects, *e.g.* in example α, where a central cyst is completely missed in the histological slide (Fig. 5**a**). The selection bias is even stronger when assessing cartilage repair results based on 2 mm diameter biopsy samples. According to the latest data of the German Cartilage Registry (Niemeyer *et al.*, 2015) the average human cartilage defect has an area of 3.68 ± 0.23 cm^2. In this case, a 2 mm biopsy would cover less than 1 % of the defect area and is assumed to be representative for this defect. Considering the large variation of repair tissue seen within this study, the authors see low predictive power in the use of biopsies in human subjects for follow-up of cartilage repair. Together with the required surgical procedure, tissue harvesting and the new focal defect that is created, we oppose this method of cartilage assessment ethically.

In contrast to histology, UBM data include the whole defect and therefore all structures contribute equally to the evaluation. Furthermore, the UBM measurements are non-destructive and therefore repeatable and need less preparation time.

We expected that the filling of the defect, expressed as *DefFill.V,* could be used as an overall parameter of the repair quality. Therefore, for each score with several items, we correlated the cumulative scores to *DefFill.V* as well as to the soft-tissue component *RepT.V*. For all scores we found highly significant, good correlations to *DefFill.V*. Moreover, a good correlation between *DefFill.V* and the ICRS-II item "overall assessment" was observed. For *RepT.V* these correlations were still significant, but only moderate. In the present study the average *DefFill.V* was 72 %, consisting of 63 % repair tissue and 9 % bone overgrowth. We hypothesise that variations in the ratio of bone overgrowth to repair tissue have caused the reduced correlation strengths for *RepT.V* to the established

classification systems. Furthermore, the established classification systems do not or only partially assess the amount of bone overgrowth. A good example for this is ε, the best outcome of the present study (Fig. 5**e**): Whereas all macroscopic and histopathological classification systems resulted in rating with the highest or near-highest scores (mean ICRS-II: 83), the "overall assessment" of ICRS-II was only scored at 60 of 100, because the rater takes the substantial amount of bone overgrowth into account more than with any other item of the classification systems. Therefore, the lower correlation strength for *RepT.V* is not surprising. However, the amount of soft repair tissue that is actually filling the defect is valuable diagnostic information. This is because the amount of bone overgrowth reduces the thickness of the repair tissue, which is then "more susceptible to damage from shear forces" (Gomoll *et al.*, 2010).

We conclude that the *DefFill.V* is the most important and robust parameter to describe the quality of the repair tissue based on UBM data. This is supported by Bonasia *et al.* who found that the item "overall assessment" of ICRS-II is the most reliable of ten major histopathological classification systems (Bonasia *et al.*, 2015). Furthermore, the comprehensive meta-analysis of Blackman *et al.* (Blackman *et al.,* 2013) revealed that the MOCART item "Degree of defect repair" shows the most significant correlations to clinical outcome measures, and according to Marlovits *et al.* (Marlovits *et al.*, 2006), the MOCART item "Degree of defect repair" also shows the best correlations to patient reported outcomes two years after autologous chondrocyte transplantation. For completeness, it should be mentioned that other studies found no correlations between MOCART "Degree of defect repair" and patient-reported outcomes (Ebert *et al.*, 2014). In future experiments, the use of UBM for the characterisation of cartilage repair tissue should focus on arthroscopic or transcutaneous application. Arthroscopic application of ultrasound in the knee to assess cartilage quality has already successfully been shown (Kaleva *et al.*, 2011; Liukkonen *et al.*, 2014). Transcutaneous quantitative ultrasound assessment of hyaline cartilage is barely investigated. However, important regions like the femoral condyles and trochlea can be visualised with commercially-available ultrasound imaging systems transcutaneously (Möller *et al.*, 2008). This would be the most patient-friendly and economical method, at least for those cartilage regions that are accessible.

When translating the present method to application in human knees, several problems need to be discussed. First of all, the average size of focal cartilage defects in human is usually one magnitude larger. Therefore, a UBM for human application needs to be scaled to acquire a larger area. This problem can overcome by the use of commercially available high-frequency ultrasound array scanners. Fully automated post-processing would be ideal, but this would need thoroughly developed software algorithms. However, first of all, this requires a standardised ultrasound imaging protocol for the investigation of articular cartilage and cartilage repair tissue. This should include definition of the anatomical sites accessible by ultrasound and the description of structures which can be seen and used as landmarks. From a technical point of view, solutions are

required to enable imaging of a 3D cartilage volume *in vivo*. Ultrasound imaging in 3D already exists in medical devices but probably needs adaptation for cartilage imaging. This could be done, *e.g.* with an ultrasound array in combination with an optical tracking system, with a customised 3D ultrasound scanner or with 2D arthroscopic ultrasound registered to a 3D MRI data set. It is important to note that transcutaneous ultrasound would require lower frequencies, which would lead to lower spatial resolutions compared to our study. However, we expect that due to the thicker cartilage in humans, the precision could be similar to the present study. Furthermore, the mathematical model to estimate the cartilage surface and cartilage-bone interface may need to be adapted to different anatomical sites and different shapes.

In summary, we have shown how 3D ultrasound biomicroscopy data can be used for volumetric quantification of cartilage repair tissue. The parameter for defect filling was significantly correlated with all established classification systems of macroscopic evaluation, MRI and histology. This method, upon translation into transcutaneous application in either human or animal studies, could considerably support focal cartilage repair follow-up and research activities. It could allow for quantification of the cartilage thickness, defect size, defect filling and malformations of the subchondral bone plate. However, in contrast to MRI and histology, ultrasound cannot visualise changes in the bone below the subchondral bone plate, *e.g.* deep cysts or oedema.

In addition to volumetric characterisation, the quality of the repair tissue can be assessed based on the ultrasound backscatter signals. As already shown, fibrotic cartilage tissue exhibits significantly more backscatter than healthy hyaline cartilage (Gelse *et al*., 2010). Therefore, repair tissue could be further differentiated into fibrotic or hyaline-like tissue based on the ultrasound backscatter analysis.

Conclusions

To conclude, UBM has been demonstrated to be suitable to evaluate cartilage repair outcomes volumetrically. The defect filling is the most valuable parameter enabling further differentiation into soft repair tissue and bone overgrowth. Moreover, a quantitative analysis of ultrasound backscatter from the cartilage matrix could be conducted to differentiate between fibrotic repair tissue and hyaline-like tissue. Further efforts should be made to combine these methods in arthroscopic or transcutaneous applications to enable for non-destructive individual follow-up. Such approach could become a useful method for characterisation and comparison of cartilage repair outcomes in clinical studies, objectively.

Acknowledgements

References

Blackman AJ, Smith MV, Flanigan DC, Matava MJ, Wright RW, Brophy RH (2013) Correlation between magnetic resonance imaging and clinical outcomes after cartilage repair surgery in the knee a systematic review and meta-analysis. Am J Sports Med **41**: 1426-1434.

Bonasia DE, Marmotti A, Massa ADF, Ferro A, Blonna D, Castoldi F, Rossi R (2015) Intra- and inter-observer reliability of ten major histological scoring systems used for the evaluation of *in vivo* cartilage repair. Knee Surg Sports Traumatol Arthrosc **23**: 2484-2493.

Buschmann MD, Saris DB (2011) Introduction to the International Cartilage Repair Society recommendation papers. Cartilage **2**: 99.

Cameron ML, Briggs KK, Steadman JR (2003) Reproducibility and reliability of the Outerbridge classification for grading chondral lesions of the knee arthroscopically. Am J Sports Med **31**: 83-86.

Cleveland WS (1979) Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc **74**: 829-836.

Ebert JR, Smith A, Fallon M, Wood DJ, Ackland TR (2014) Correlation between clinical and radiological outcomes after matrix-induced autologous chondrocyte implantation in the femoral condyles. Am J Sports Med **42**: 1857-1864.

Gelse K, Olk A, Eichhorn S, Swoboda B, Schoene M, Raum K (2010) Quantitative ultrasound biomicroscopy for the analysis of healthy and repair cartilage tissue. Eur Cell Mater **19**: 58-71.

Goebel L, Zurakowski D, Müller A, Pape D, Cucchiarini M, Madry H (2014) 2D and 3D MOCART scoring systems assessed by 9.4 T high-field MRI correlate with elementary and complex histological scoring systems in a translational model of osteochondral repair. Osteoarthritis and Cartilage **22**: 1386-1395.

Gomoll AH, Madry H, Knutsen G, van Dijk N, Seil R, Brittberg M, Kon E (2010) The subchondral bone in articular cartilage repair: current problems in the surgical management. Knee Surg Sports Traumatol Arthrosc **18**: 434-447.

Hepp P, Osterhoff G, Niederhagen M, Marquass B, Aigner T, Bader A, Josten C, Schulz R (2009) Perilesional changes of focal osteochondral defects in an ovine model and their relevance to human osteochondral injuries. J Bone Joint Surg Br **91-B**: 1110-1119.

Hunziker EB, Lippuner K, Keel MJB, Shintani N (2015) An educational review of cartilage repair: precepts & practice – myths & misconceptions – progress & prospects. Osteoarthritis Cartilage **23**: 334-350.

Hurtig MB, Buschmann MD, Fortier LA, Hoemann CD, Hunziker EB, Jurvelin JS, Mainil-Varlet P, McIlwraith CW, Sah RL, Whiteside RA (2011) Preclinical studies for

cartilage repair: recommendations from the International Cartilage Repair Society. Cartilage **2**: 137-152.

Irrgang JJ, Anderson AF, Boland AL, Harner CD, Kurosaka M, Neyret P, Richmond JC, Shelborne KD (2001) Development and validation of the international knee documentation committee subjective knee form. Am J Sports Med **29**: 600-613.

Jackson DW, Lalor PA, Aberman HM, Simon TM (2001) Spontaneous repair of full-thickness defects of articular cartilage in a goat model A Preliminary study. J Bone Joint Surg Am **83**: 53-64.

Jurvelin JS, Räsänen T, Kolmonens P, Lyyra T (1995) Comparison of optical, needle probe and ultrasonic techniques for the measurement of articular cartilage thickness. J Biomech **28**: 231-235.

Kaleva E, Virén T, Saarakkala S, Sahlman J, Sirola J, Puhakka J, Paatela T, Kröger H, Kiviranta I, Jurvelin JS, Töyräs J (2011) Arthroscopic ultrasound assessment of articular cartilage in the human knee joint. Cartilage **2**: 246-253.

Liukkonen J, Lehenkari P, Hirvasniemi J, Joukainen A, Virén T, Saarakkala S, Nieminen MT, Jurvelin JS, Töyräs J (2014) Ultrasound arthroscopy of human knee cartilage and subchondral bone *in vivo*. Ultrasound Med Biol **40**: 2039-2047.

Mainil-Varlet P, Aigner T, Brittberg M, Bullough P, Hollander A, Hunziker E, Kandel R, Nehrer S, Pritzker K, Roberts S, Stauffer E (2003) Histological assessment of cartilage repair: Aareport by the Histology Endpoint Committee of the International Cartilage Repair Society (ICRS). J Bone JointSurg Am **85**: 45-57.

Mainil-Varlet P, Damme BV, Nesic D, Knutsen G, Kandel R, Roberts S (2010) A new histology scoring system for the assessment of the quality of human cartilage repair: ICRS II. Am. J Sports Med **38**: 880-890.

Marlovits S, Singer P, Zeller P, Mandl I, Haller J, Trattnig S (2006) Magnetic resonance observation of cartilage repair tissue (MOCART) for the evaluation of autologous chondrocyte transplantation: determination of interobserver variability and correlation to clinical outcome after 2 years. Eur J Radiol **57**: 16-23.

Marquass B, Schulz R, Hepp P, Zscharnack M, Aigner T, Schmidt S, Stein F, Richter R, Osterhoff G, Aust G, Josten C, Bader A (2011) Matrix-associated implantation of predifferentiated mesenchymal stem cells *versus* articular chondrocytes *in vivo* results of cartilage repair after 1 year. Am J Sports Med **39**: 1401-1412.

Möller I, Bong D, Naredo E, Filippucci E, Carrasco I, Moragues C, Iagnocco A (2008) Ultrasound in the study and monitoring of osteoarthritis. Osteoarthritis Cartilage **16: Suppl 3** doi:10.1016/j.joca.2008.06.005.

Niemeyer P, Schweigler K, Grotejohann B, Maurer J, Angele P, Aurich M, Becher C, Fay J, Feil R, Fickert S, Fritz J, Hoburg A, Kreuz P, Kolombe T, Laskowski J, Lützner J, Marlovits S, Müller PE, Niethammer T, Pietschmann M, Ruhnau K, Spahn G, Tischer T, Zinser W, Albrecht D (2015) [The German Cartilage Registry (KnorpelRegister DGOU) for evaluation of surgical treatment for cartilage defects: experience after six months including first demographic data]. Z Orthop Unfall **153**: 67-74.

O'Driscoll SW, Keeley FW, Salter RB (1988) Durability of regenerated articular cartilage produced by free autogenous periosteal grafts in major full-thickness defects in joint surfaces under the influence of continuous passive motion. A follow-up report at one year. J Bone Joint Surg Am **70**: 595-606.

Outerbridge RE (1961) The etiology of chondromalacia patellae. J Bone Joint Surg Br **43-B**: 752-757.

Parfitt AM, Drezner MK, Glorieux FH, Kanis JA, Malluche H, Meunier PJ, Ott SM, Recker RR (1987) Bone histomorphometry: Standardization of nomenclature, symbols, and units: Report of the ASBMR histomorphometry nomenclature committee. J Bone Miner Res **2**: 595-610.

Raum K, Obrien WD (1997) Pulse-echo field distribution measurement technique for high-frequency ultrasound sources. IEEE Trans. Ultrason Ferroelectr Freq Control **44**: 810-815.

Raum K, Ozguler A, Morris SA, O'Brien Jr WD (1998) Channel defect detection in food packages using integrated backscatter ultrasound imaging. IEEE Trans. Ultrason Ferroelectr Freq Control **45**: 30-40.

Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD (1998) Knee Injury and Osteoarthritis Outcome Score (KOOS) – development of a self-administered outcome measure. J Orthop Sports Phys Ther **28**: 88-96.

Rutgers M, van Pelt MJ, Dhert WJ, Creemers LB, Saris DB (2010) Evaluation of histological scoring systems for tissue-engineered, repaired and osteoarthritic cartilage. Osteoarthritis Cartilage **18**: 12-23.

Schöne M, Männicke N, Gottwald M, Göbel F, Raum K (2013) 3-D high-frequency ultrasound improves the estimation of surface properties in degenerated cartilage. Ultrasound Med Biol **39**: 834-844.

Töyräs J, Laasanen MS, Saarakkala S, Lammi MJ, Rieppo J, Kurkijärvi J, Lappalainen R, Jurvelin JS (2003) Speed of sound in normal and degenerated bovine articular cartilage. Ultrasound Med Biol **29**: 447-454.

Trattnig S, Ba-Ssalamah A, Pinker K, Plank C, Vecsei V, Marlovits S (2005) Matrix-based autologous chondrocyte implantation for cartilage repair: noninvasive monitoring by high-resolution magnetic resonance imaging. Magn. Reson. Imaging **23**: 779-787. doi:10.1016/j.mri.2005.04.010.

Willbold E, Witte F (2010) Histology and research at the hard tissue-implant interface using Technovit 9100 New embedding technique. Acta Biomater **6**: 4447-4455.

Windt TS de, Welsch GH, Brittberg M, Vonk LA, Marlovits S, Trattnig S, Saris DBF (2013a) Is magnetic resonance imaging reliable in predicting clinical outcome after articular cartilage repair of the knee? A systematic review and meta-analysis. Am J Sports Med **41**: 1695-1702.

Windt TS de, Welsch GH, Brittberg M, Vonk L, Marlovits S, Trattnig S, Saris DB, Blackman AJ, Smith MV, Flanigan DC, Matava MJ, Wright RW, Brophy RH (2013b) Correlation between magnetic resonance imaging and clinical outcomes after knee cartilage repair: letter to the editor. Am J Sports Med **41**: doi: 10.1177/0363546513510140.

Zscharnack M, Hepp P, Richter R, Aigner T, Schulz R, Somerson J, Josten C, Bader A, Marquass B (2010) Repair of chronic osteochondral defects using predifferentiated

mesenchymal stem cells in an ovine model. Am J Sports Med **38**: 1857-1869.

Web References

1. Draft guidance for industry: preparation of IDEs and INDs for products intended to repair or replace knee cartilage, U.S. Food and Drug Administration. http://www.fda.gov/BiologicsBloodVaccines/ GuidanceComplianceRegulatoryInformation/Guidances/ CellularandGeneTherapy/ucm072952.htm [21-05-2015].

2. Reflection paper on *in-vitro* cultured chondrocyte containing products for cartilage repair of the knee (2010). European Medicines Agency – Committee For Advanced Therapies. http://www.ema.europa.eu/docs/ en_GB/document_library/Scientific_guideline/2010/05/ WC500090887.pdf [21-05-2015].

Editor's Note: All questions/comments by the reviewers were answered by text changes. There is hence no Discussion with Reviewers section. Scientific Editor in charge of the paper: Mauro Alini.