



Analysis

Histopathological Grading of Oral Mucosal Chronic Graft-versus-Host Disease: Large Cohort Analysis



Victor Tollemar¹, Nikcole Tudzarovski¹, Gunnar Warfvinge², Naom Yarom^{3,4}, Mats Remberger^{5,6}, Robert Heymann^{1,7}, Karin Garming Legert¹, Rachael V. Sugars^{1,*}

¹ Division of Oral Diagnostics and Rehabilitation, Department of Dental Medicine, Karolinska Institutet, Stockholm, Sweden.

² Department of Oral Pathology, Faculty of Odontology, Malmö University, Malmö, Sweden

³ Oral Medicine Unit, Sheba Medical Center, Tel-Hashomer, Israel

⁴ School of Dental Medicine, Tel-Aviv University, Tel-Aviv, Israel

⁵ Clinical Research and Development Unit, Uppsala University Hospital, Uppsala, Sweden

⁶ Institution for Medical Science, Uppsala University, Uppsala, Sweden

⁷ Trauma and Regenerative Medicine, Craniofacial Diseases, Karolinska University Hospital, Stockholm, Sweden

Article history:

Received 8 April 2020

Accepted 28 June 2020

Keywords:

Oral cGVHD

Histopathology

Grading

Large cohort

Hematopoietic cell transplantation

A B S T R A C T

Graft-versus-host disease (GVHD) can manifest as acute or chronic complications in patients after hematopoietic cell transplantation (HCT). Oral chronic GVHD (cGVHD) occurs in approximately 70% of HCT recipients and includes lichenoid-like mucosal reactions, restricted mouth opening, and salivary gland dysfunction. However, the underlying histopathological presentation remains to be validated in large cohorts. We characterized the histopathological features of oral mucosal cGVHD and devised a scoring model in a large patient cohort (n = 112). Oral mucosal biopsy sections (n = 303) with and without oral cGVHD were identified from archived and current HCT recipients with additional healthy controls. Histological screening was performed on hematoxylin and eosin-stained and periodic acid-Schiff-stained sections. A points-based grading tool (0 to 19, grade 0 to IV) was established based on intraepithelial lymphocytes and band-like inflammatory infiltrate, atrophic epithelium with basal cell liquefaction degeneration, including apoptosis, as well as separation of epithelium and pseudo-rete ridges. Validation involved 62 biopsy specimens, including post-HCT (n = 47) and healthy (n = 15) specimens. Remaining biopsy specimens (n = 199) were blindly graded by 3 observers. Histological severity was correlated with clinical diagnostic and distinctive features, demonstrating a spectrum of individual patient severity, including frequent signs of subclinical GVHD in healthy mucosa. However, oral cGVHD presented with significantly higher (P < .001) scores compared with HCT controls, with moderate to high positive likelihood ratios for inflammatory infiltrate, exocytosis, and basal membrane alterations. The grade II-IV biopsy specimens demonstrated a histopathological diagnosis of active mucosal lichenoid-like cGVHD, highlighting the importance of correlating clinical presentation with the dynamic histopathological processes for improved patient stratification. In addition, this tool could be used for assessing treatments, pathological processes, and immune cellular content to provide further insight into this debilitating disease.

© 2020 American Society for Transplantation and Cellular Therapy. Published by Elsevier Inc. This is an open access article under the CC BY license. (<http://creativecommons.org/licenses/by/4.0/>)

INTRODUCTION

Allogeneic hematopoietic cell transplantation (HCT) is used worldwide for various conditions, mainly to treat leukemias [1]. Improved transplantation technologies have facilitated the increased use of HCT, which in turn has led to an increased incidence of chronic graft-versus-host disease (cGVHD). However, this rise has been shown to be independent of associated

risk factors, such as enrollment of older patients, augmented use of peripheral blood stem cells, use of unrelated donors, and use of different HCT strategies [2]. The leading cause of death is relapse of the initial disease, followed by infection and organ failure; cGVHD is often considered the major cause of non-relapse-related mortality, with reduced quality of life and disability [2,3].

The precise mechanisms by which GVHD starts and progresses remain vague, although donor T and B cell involvement have been demonstrated to initiate tissue injury [4]. cGVHD resembles autoimmune-like disorders, with signs and symptoms that affect multiple organs and sites, and it has

Financial disclosure: See Acknowledgments on page 1978.

*Correspondence and reprint requests: R.V. Sugars, Division of Oral Diagnostics and Rehabilitation, Department of Dental Medicine, Karolinska Institutet, Alfred Nobel's Väg 8, 14104 Huddinge, Stockholm, Sweden.

E-mail address: Rachael.Sugars@ki.se (R.V. Sugars).

<https://doi.org/10.1016/j.bbmt.2020.06.031>

1083-8791/© 2020 American Society for Transplantation and Cellular Therapy. Published by Elsevier Inc. This is an open access article under the CC BY license. (<http://creativecommons.org/licenses/by/4.0/>)

been stratified into organ-specific diagnostic and distinctive clinical signs by the National Institutes of Health (NIH) Consensus Development Project [5,6]. cGVHD in the oral cavity presents in approximately 70% of allogeneic HCT recipients and is considered one of most prevalent post-transplantation complications and, along with skin involvement, one of the earliest signs of cGVHD [3,7,8]. The most recent diagnostic criteria for oral cGVHD are oral lichen planus (OLP)-like lesions, and distinctive signs include xerostomia, mucocoeles, mucosal atrophy, ulcers, and pseudomembranes [6]. Distinctive signs require confirmation to rule out viral or bacterial infections, together with histopathological analysis to confirm the minimum diagnostic score of “likely” GVHD [6,9]. Oral mucosal presentation, salivary gland dysfunction, and limited mouth opening have been proposed as 3 different disease presentations and should be considered individually [10].

Ancillary to clinical examination are histological criteria for diagnosis confirmation. The 2015 NIH Consensus Project refined the minimum histological criteria for oral mucosal cGVHD to include “lichenoid interface lymphocytes with infiltration of the mucosa (exocytosis) and variable apoptosis” [9]. Numerous histological observations in the oral mucosa have been described in HCT recipients, including subepithelial inflammation [11–18], acanthosis [17], atrophy/flattening of rete ridges [11,12], intraepithelial inflammation/exocytosis [15,17–19], epithelial/hydropsic degeneration/necrosis including vacuolization [11,12,14–16,19,20], apoptosis including satellitosis [12,15–18,20], basal membrane thickening/clefting [15,17,19], and keratinization [16,18]. Most studies have been performed with small groups of patients and characterized according to histopathological strategies, but with the minimal classification published by the NIH Consensus Project, studies have reverted to using those inclusion criteria [9,12,21]. The NIH criteria collate all oral cGVHD patients into 1 consolidated group, but there is a need to provide guidance toward the final diagnosis of “possible” or “likely” GVHD. Individual patient responses are diverse, and in light of the proposed 3 separate disease states, the research findings are difficult to interpret and compare with only minimal criteria. Early studies by Sale et al [12] suggested 2 grades of oral histopathological severity based on inflammation and epithelial necrosis. Thereafter, only 1 original report described an oral GVHD grading scheme, published by Horn et al [19]. Both studies were designed as a combined mucosal and salivary gland evaluation scheme focusing on subjective definitive features to derive specific grades; however, they were conducted more than 25 years ago, before the NIH Consensus Project.

Thus, there is a significant need to validate the NIH histopathology consultation form (<https://www.astct.org/practice-resources/nih-chronic-gvhd-consensus-project>) and define histopathological guidelines in a sufficiently large cohort to stratify the histological presentations and facilitate clinical guidance in research trials [21]. In a large cohort, we have validated and specified the NIH grading scheme to assess the specific features individually to reach an objective final grade for oral mucosal cGVHD histopathology. The grading scheme is designed to facilitate pathologists’ working practices, leading to a better understanding of the biology and providing meaningful information to clinicians.

METHODS

Ethical Permissions

All specimen collections and assessments were performed in accordance with the Declaration of Helsinki and with approval from the Stockholm Local Ethics Board (registration nos. 2013-1241-31-1 and 2014/1184-31-1).

Biopsy Collection

Biopsy specimens were retrieved either retrospectively from Stockholm’s Medical Biobank (SMB) or obtained prospectively (after 2013) from allogeneic HCT recipients referred to the Oral and Maxillofacial Surgery Clinic at Karolinska University Hospital. Archived SMB cases (1977 to 2011) for patients with hematologic conditions were reviewed ($n = 789$ patient records) and included based on allogeneic HCT with clinical information covering late phase (> 100 days) post-HCT (Figure 1A). These patients followed a clinical protocol with observations performed every 3 months post-HCT and routine biopsy specimens of labial mucosa obtained at multiple time points irrespective of oral status. The use of specimens from different times during the study period allowed consideration of all phases of cGVHD pathobiology. Patients from 2013 were referred routinely at 6 months post-HCT and at additional later time points for longstanding cGVHD complications. At consultation (K.G.L), voluntary patients were clinically scored according to a modified oral mucosal rating scale (OMRS), and a 5-mm buccal mucosa punch biopsy was obtained in a nonulcerated area close to the second molar ($n = 16$) [22]. Healthy volunteers ($n = 15$; 13 males and 2 females), with a median age of 37.7 years (range, 17 to 76 years) underwent punch biopsy to obtain a 5-mm specimen at a similar site as the HCT recipients when attending the Oral and Maxillofacial Surgery Clinic at Karolinska University Hospital for oral mucosal surgery.

Retrospective Clinical Diagnostics

To align all patient biopsy specimens according to the NIH Consensus Clinical Scoring, retrospective assessment of the SMB medical records was reviewed independently by 2 oral medicine specialists from 2 different GVHD centers (K.G.L and N.Y) [6]. The biopsy specimens were assessed based on intraoral photographs and/or clinical description of mucosal status and grouped either as defined oral cGVHD, when current NIH diagnostic criteria were met, or as possible oral cGVHD, with distinctive criteria but inconclusive information to definitively rule out other possible disorders [6]. Oral HCT control specimens were identified and included as patients with normal-appearing mucosa, without clinical involvement of oral mucosal GVHD or any other mucosal condition that could be noted at the time of biopsy. The prospective cohort was also allocated into the study’s clinical groups (described above) based on NIH diagnostic criteria but with guidance using the OMRS score at time of clinical examination. When significant differences of opinion arose due to insufficient information or proposed exclusion of patients from the study, both oral medicine specialists reviewed these cases together to reach a consensus. The characteristics of the HCT patient cohort are shown in Table 1.

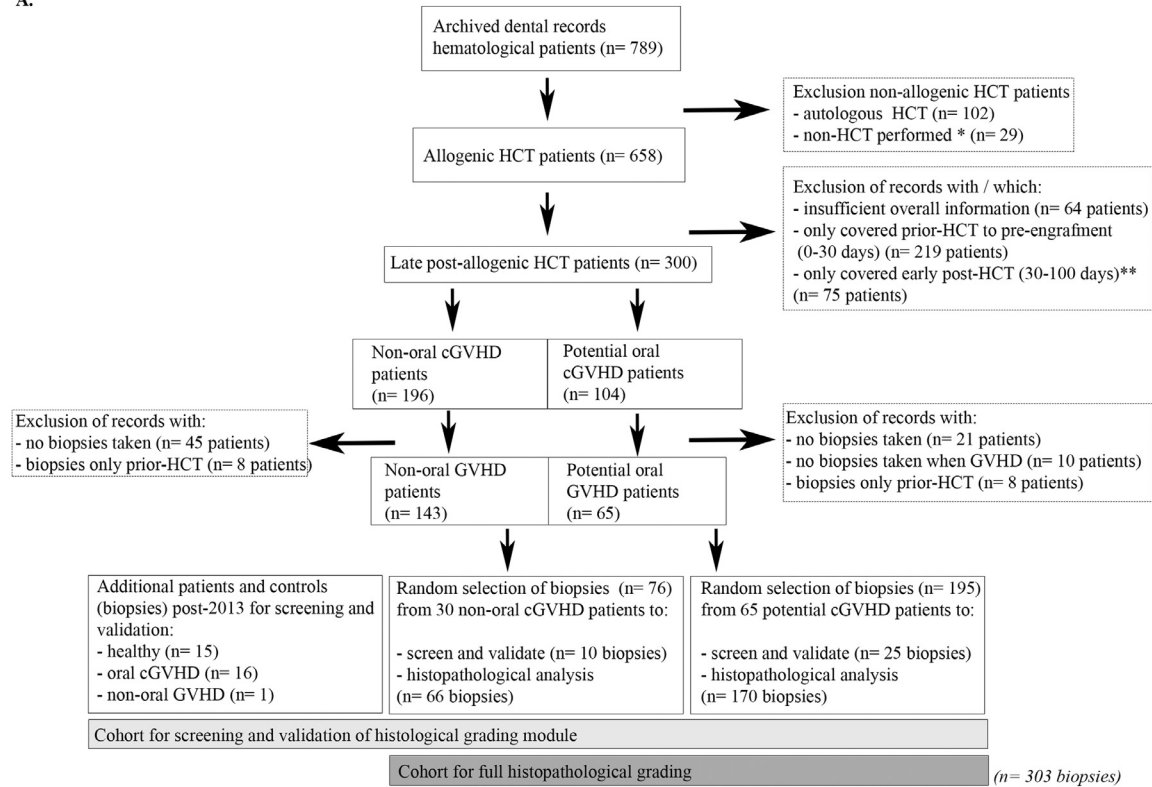
Histological Preparation and Staining

Archived biopsy specimens and/or original pathology slides (hematoxylin and eosin [HTX] stained) for patients meeting the inclusion criteria were retrieved retrospectively from SMB (Figure 1A). Additional routine staining with Mayer’s HTX (HISTOLAB Products, Gothenburg, Sweden) and periodic acid-Schiff (PAS; Merck KGaA, Darmstadt, Germany) was performed, if possible, to allow examination deeper into the tissue. PAS was included to visualize the basal membrane and any potential *Candida* spp infections. To validate the cytomegalovirus (CMV) registry data, 35 randomly selected biopsy specimens from HCT recipients ($n = 16$ CMV-negative, $n = 19$ CMV-positive) and 10 healthy volunteers were screened for the presence of CMV and compared against CMV-positive controls (Cell Marque, Rocklin, CA). Whole-slide imaging was used to digitalize all sections using a 3D midi scanner system (3DHIS-TECH, Budapest, Hungary), and all further observations were blindly viewed with Panoramic Viewer 1.15 software (HISTOLAB Products).

Histopathological Grading Scheme: Validation and Full Cohort Analyses

HTX- and PAS-stained sections from normal healthy oral mucosa, prospective biopsy specimens obtained after 2013, and 35 randomly chosen tissue specimens from SMB were used as a validation cohort to ensure a representative cross-section of all clinical presentations when setting up the histological grading tool (Figure 1A). Screening for biopsy suitability excluded 5 specimens that either could not be traced within SMB or had minimal or missing epithelium (Figure 1B). PAS staining was not possible in 4 specimens. Four assessors (V.T., N.T., G.W., and R.S.) individually screened the validation cohort blindly to investigate and identify recurring prominent oral histological features in the HTX- and PAS-stained tissues. With reference to published findings, description of features with consequent boundaries were discussed in pathology validation meetings, leading to a finalized grading scheme (Table 2) and consensus scores for the validation specimens. Features were inflammatory infiltrate and exocytosis, basal membrane changes and liquefaction degeneration, apoptosis, and atrophy of the epithelia (Table 2 and Supplementary Table S1). All features were weighted according to histopathological severity and summarized with a maximum score of 19. Cluster analysis of the validation cohort scores defined the separation of the histological grading into grades (G) 0 to GIV (G0, 0 to 2; GI, 3 to 5; GII, 6 to 9; GIII, 10 to 13 and GIV, 14 to 19) (Supplementary Figure S1). The remaining SMB cases

A.



B.

Screening and validation of the histopathological grading module (n=54 patients)	Full histopathological grading module (n= 90 patients)
Exclusion (n= 2 patients): - not detectable in the biobank (n= 2 biopsies) - insufficient quality to grade (n= 3 biopsies)***	Exclusion (n= 6 patients): - not detectable in the biobank (n= 21 biopsies) - insufficient quality to grade (n= 16 biopsies)***
Inclusion: 29 potential oral cGVHD patients - 0 = prior-HCT biopsies 0 / 0 - 5 = early post-HCT biopsies 5 / 4 - 31 = late post-HCT biopsies 31 / 30 8 non-oral cGVHD patients - 0 = prior-HCT biopsies 0 / 0 - 1 = early post-HCT biopsies 1 / 1 - 10 = late post-HCT biopsies 10 / 8 15 healthy controls biopsies 15 / 15 <i>n = 62 biopsies from 52 patients</i>	Inclusion: 57 potential oral cGVHD patients - 20 = prior-HCT biopsies 20 / 14 - 31 = early post-HCT biopsies 31 / 24 - 91 = late post-HCT biopsies 91 / 67 27 non-oral cGVHD patients - 6 = prior-HCT biopsies 6 / 3 - 16 = early post-HCT biopsies 16 / 11 - 35 = late post-HCT biopsies 35 / 29 <i>n = 199 biopsies from 84 patients</i>

C.

Complete cohort following cGVHD clinical scoring of archived HCT patients (n= 303 biopsies)
Exclusion: - during screening and validation or histological grading (n= 42 biopsies) - after clinical grading (n= 49 biopsies)
Inclusion: - defined oral cGVHD (n= 78 biopsies) - possible oral cGVHD (n= 44 biopsies) - oral HCT control (inc. no oral cGVHD and prior to oral cGVHD (n= 49 biopsies) - prior to HCT (n= 26 biopsies) - healthy (n= 15 biopsies) <i>n = 212 biopsies</i>

Figure 1. (A) A total of 789 archived (SMB) clinical records were examined from patients referred as a result of hematologic conditions to the Oral and Maxillofacial Surgery Clinic at Karolinska University Hospital. Patients were excluded and included based on various criteria and classified as clinically nonoral cGVHD or potential oral cGVHD. These patients were allocated into a cohort for screening and validation of the histological grading tool and/or a cohort for complete analysis of histological presentation in oral mucosal tissues. *Includes patients examined for HCT but with treatment plans that changed, as well as those patients who did not survive to the HCT, **Three patients of potential GVHD condition in the early phase (<100 days) were included even though not they were followed into the late phase (>100 days). (B) From those selected into the respective groupings, the total number of biopsy specimens assessed by either HTX or PAS staining for each cohort are defined. ***Nonqualitative specimens including only minor salivary glands, specimens with indistinct morphology or specimens lacking mucous membrane tissue. ****Specimens did not meet the quality criteria after further sectioning. (C) Patients meeting the criteria (n = 303 biopsy specimens) were subjected to clinical scoring for healthy, prior to HCT, HCT control (including no oral cGVHD and prior to oral cGVHD development), possible oral GVHD, and defined oral cGVHD. Archived SMB patients were retrospectively scored with the National Institutes of Health Consensus Clinical Scoring for oral cGVHD or patients from 2013 onward were diagnosed using the OMRS and considered as defined oral cGVHD. Biopsy specimens were excluded if insufficient information was provided. Thus, the final cohort comprised 212 biopsy specimens for analysis.

Table 1
Clinical characteristics of HCT-patients.

Characteristics	n (% or range)
No. of patients	112
No. of biopsies	288
Age and gender	
Median age (y) at HCT	32.7 (1-67)
Children (<18 years)	22 (19.6)
Ratio male/female/unknown	80/31/1 (71.4/27.6/0.9)
Year of HCT	
-1991	73 (65.2)
1992-2000	23 (20.5)
2001-2009	0
2010-	16 (14.3)
Donor	
HLA identical related	92 (82.1)
MUD	12 (10.7)
Other (MM related, MRD)	7 (6.2)
Unknown	1 (1.0)
Gender (unmatched/matched/ unknown)	54/57/1 (48.2/50.8/1.0)
CMV match (pos / neg / miss / unknown)	52/22/33/5 (46.4/19.6/29.5/4.5)
Disease	
Chronic Leukemia	33 (29.4)
Acute Leukemia	47 (42.0)
AA	11 (9.8)
MM	6 (5.4)
Lymphoma	6 (5.4)
Other/unknown (MDS/MPS, Metab, MF)	9 (8.0)
Conditioning regimen	
Myeloablative	88 (78.5)
Reduced intensity	23 (20.5)
Unknown	1 (1.0)
GVHD prophylaxis	
CsA	15 (13.4)
MTX	20 (17.9)
CsA + MTX	63 (56.31)
Other/unknown, (Tac + Sir, TcD)	14 (12.5)
HCT source	
BM/PBSC/Unknown	89/22/1 (79.4/19.6/1.0)
CMV infection	
(pos / neg / unknown)	48/63/1 (42.8/56.2/1.0)
aGVHD grade	
0	32 (28.5)
1	57 (50.8)
2	20 (17.8)
3	1 (1.0)
4	1 (1.0)
Unknown	1 (1.0)
cGVHD score	
None	35 (31.2)
Mild	51 (45.5)
Moderate	16 (14.2)
Severe	9 (8.0)
Unknown	1 (1.0)

MUD – matched unrelated donor, MM - multiple myeloma, MRD – matched related donor, MDS – myelodysplastic syndrome, MPS – myeloproliferative syndrome, Metab – metabolic disorders, CsA – cyclosporine A, MTX – methotrexate, Tac – tacrolimus, Sir – sirolimus, TcD – T-cell depletion, BM – bone marrow, PBSC – peripheral blood stem cell.

(n = 236 biopsy specimens) were assessed for the full histopathological analysis or were excluded if they did not meet the quality criteria (n = 37) (Figure 1B). Thus, 199 biopsy specimens (HTX and PAS) were individually blindly graded by 3 assessors (V.T., N.T., and R.S.). All sections were given a severity score between 0 and 19 (Table 2, Supplementary Figure S1, and Supplementary Table S1).

Statistics and Robustness Testing

Jenks natural breaks optimization for 1-dimensional data was used to define boundary values and to classify the data into groups [23,24]. Cohen's kappa with quadratic weighting was performed to test the agreement by chance of the histopathological grading and for the clinical classification of the 2 oral medicine specialists [25]. Statistical tests were performed using Prism 8 (GraphPad Software, La Jolla, CA), including Student's t-test and the Kruskal-Wallis test with Dunn's correction for multiple comparisons to assess the histological groups and severity scores. Receiver operator characteristic (ROC) curve plots were determined for each histological feature related to oral clinical presentation of prior HCT, oral HCT control, possible cGVHD, and defined cGVHD. The 95% confidence intervals (CIs) were set using the Wilson-Brown method with area under the curve (AUC) and likelihood ratio (LR) calculated. Significance was considered at a P value <0.05.

RESULTS

Patient Characteristics

We identified 303 oral specimens from 112 HCT recipients and 15 healthy controls for the purpose of developing the grading module and subsequent analysis (Figure 1B). The HCT patients included are summarized in Table 1. The majority (n = 80; 71.3%) underwent transplantation due to leukemia, and the overall cGVHD peak severity was mainly mild (n = 51; 45.5%). The median time of cGVHD diagnosis was 140 days post-HCT. One patient developed secondary cancer in the tongue mucosa. Of the 48 patients documented with CMV infection, 45 were diagnosed before day 100 (median, 45 days), and 1 had an unknown date of onset. All except 2 biopsy specimens included from CMV-positive patients were taken at least 30 days post-CMV diagnosis (days 1 and 29).

Retrospective Clinical Diagnosis of Oral cGVHD

Patient records from the archived SMB were clinically screened by 2 oral medicine specialists to allocate each biopsy a grade of prior HCT, oral HCT control, possible oral cGVHD, or defined oral cGVHD to follow the NIH clinical criteria for oral cGVHD (Figure 1C) [6]. From the original collection of 303 biopsy specimens, 15 were from healthy controls, 42 were excluded following the initial histopathological phase (detailed below), and 16 obtained after 2013 were from patients who had been diagnosed at their clinical visits, leaving a total of 230 specimens for clinical evaluation. An initial screening of records was performed to validate the clinical information for each biopsy included, by 1 of the specialists, resulting in exclusion of 31 specimens from the individual assessment. The remaining 199 graded biopsy specimens showed intraobserver concordance between the oral medicine specialists of 0.8125 (Supplementary Figure S2). A second phase of joint clinical evaluation was held to reach consensus for the 31 specimens presenting with limited information and those not in agreement (n = 66). Thirty-six specimens were also included from the expert consensus conference. Thus, the complete cohort was divided into healthy (normal oral mucosa; n = 15), prior to HCT (n = 26), oral HCT control (n = 49; subdivided into no oral GVHD [n = 34] and prior to oral cGVHD [n = 15]), possible oral cGVHD (n = 44), and defined oral cGVHD (n = 78) (Figure 1C).

Table 2

Histological grading module for defining features of oral mucosal cGVHD with scored points. Assessment of 212 biopsies that included 197 biopsies from HCT patients and 15 healthy control biopsies.

	Grade				
Features	0	1	2	3	4
1 Inflammation infiltrate	Small numbers of scattered cells, papilla focused	Sparsely clustered cells	Tightly clustered cells	Band-like infiltrate	Extensive band-like infiltrate
Score:	0	1	2	3	4
n (%)	71 (33.5)	70 (33.0)	40 (18.9)	22 (10.4)	9 (4.2)
2 Intra-epithelial infiltration of lymphocytes	None / occasional	Sporadic	Focal	Widespread	
Score:	0	1	2	3	
n (%)	56 (26.4)	81 (38.2)	40 (18.9)	35 (16.5)	
3 Liquefaction degeneration	None	Sporadic	Widespread	Confluent	
Score:	0	1	2	3	
n (%)	77 (36.3)	63 (29.7)	31 (14.6)	41 (19.3)	
4 Apoptosis	None / occasional	Sporadic	Widespread		
Score:	0	1	2		
n (%)	117 (55.2)	64 (30.2)	31 (14.6)		
5 Basal membrane alteration	Thin intact	Increased thickness			Thinning / loss with detachment, and / or pseudo rete ridges
Score:	0	1			4
n (%)	115 (54.2)	72 (34.0)			25 (11.8)
6 Flattening / atrophy	Normal rete ridges	Some flattening of rete ridges across biopsy (<25%)	Flattening of rete ridges across biopsy (25-75%)	Flat atrophic oral epithelia across biopsy (75-100%)	
Score:	0	1	2	3	
n (%)	40 (18.9)	72 (34.0)	55 (25.9)	45 (21.2)	
					Total score: 19

Grade 0 - 0-2; Grade I - 3-5; Grade II - 6-9; Grade III - 10-13; Grade IV - 15-19.

Establishment of the Histopathological Grading Scheme

The histopathological features were stratified into 6 major groups that changed with severity, and these were compiled into the histopathological grading scheme with associated scoring using a random selection of biopsy specimens for validation purposes (n = 62) (Table 2, Figures 1A and 2, and Supplementary Table S1). Inflammatory infiltrate was graded based on the resemblance to an OLP-band like clustering of cells (score 0 to 4), and intraepithelial infiltration of lymphocytes was scored between 0 and 3 depending on occurrence. Basal cell degeneration was divided into liquefaction degeneration, where epithelial cells became necrotic with vacuolization/spongiosis (score 0 to 3), basal membrane alterations with changes in basal membrane thickness that eventually led to Max-Joseph spaces or pseudo-rete ridges in the most severe cases (score 0 to 4), and occurrence of apoptosis (scored 0 to 2). The final feature was flattening of rete ridges with epithelial atrophy (scored 0 to 3). A weighted score was assigned for each group of categorical features, with a maximum score of 19 across the complete grading scheme. Clustering of the 62 specimens (Supplementary Figure S1) in the validation cohort were used to identify the severity groups.

Histopathological Grading Analysis of the Whole Cohort

Biopsy specimens from the remaining archived SMB cohort were blindly graded by 3 assessors (V.T., N.T., and R.S.), with inter-observer concordance of 0.7076, 0.7756, and 0.8425, respectively (Supplementary Figure S3) [25]. The combined grading of the validation cohort with the remaining large group (n = 212) showed that each histological grade was well represented (Supplementary Figure S4A). Within the assessment of 212 specimens (Table 2), 66.5% presented with some degree of inflammation but only 14.6% exhibited OLP-band-like pattern. More than one-third (35.4%) showed focal or widespread lymphocyte exocytosis, and some degree of liquefaction degeneration was observed across most specimens (63.7%). More than one-half did not exhibit any apoptosis or basal membrane alterations, although the proportion presenting with the most severe scores was relatively high. Furthermore, some degree of flattening to complete atrophy of the epithelium was observed in more than 80% of biopsy specimens.

Correlation of Histopathological Findings to Oral Clinical Assessments

Across the different oral clinical classifications, a broad spectrum of histological scores was commonplace (Figure 3A). Fifteen normal healthy controls exhibited few if any

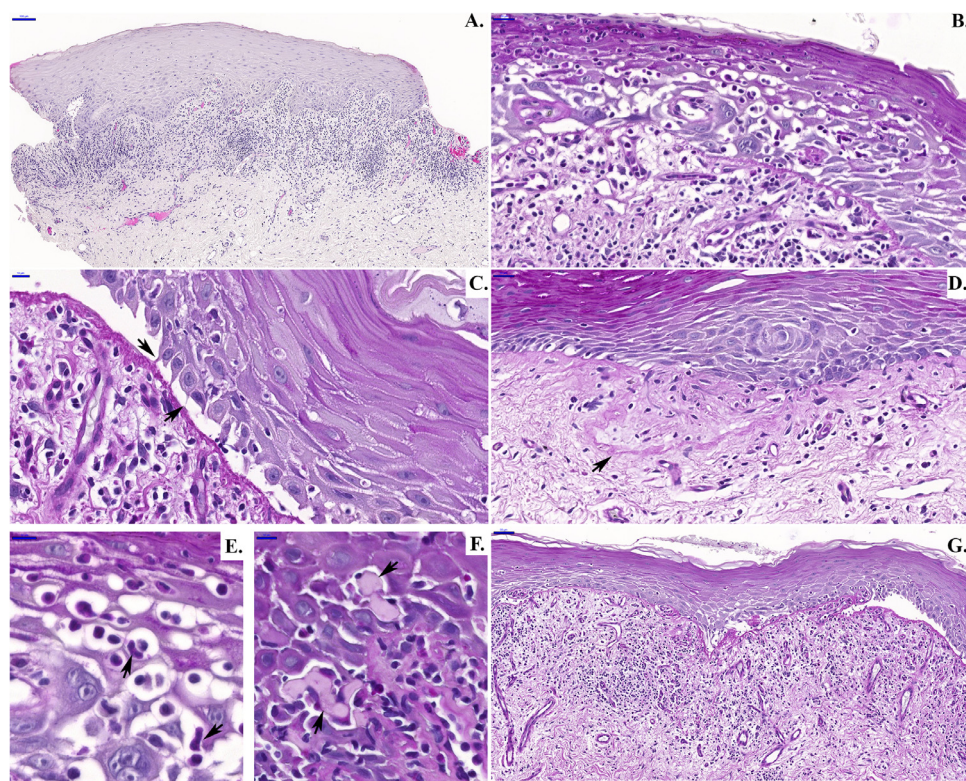


Figure 2. Representative images from the most severe grades of features included within the histological grading module. (A) HTX staining demonstrated extensive band-like inflammation infiltrate across the whole biopsy. (Scale bar: 100 μm .) (B) Widespread intraepithelial infiltration of lymphocytes and liquefaction degeneration of the epithelial cell layer stained with PAS. (Scale bar: 20 μm .) (C) Basal membrane alterations with thinning and loss of attachment (arrows) to basal membrane; PAS stained. (Scale bar: 10 μm .) (D) Associated with basal membrane alterations, development of pseudo-retic ridges into the underlying lamina propria could be detected with PAS staining in the most severe grades (arrow). (Scale bar: 10 μm .) (E) Apoptotic cells (arrows) were identified extensively throughout the tissues with eosinophilic bodies. (Scale bar: 10 μm .) (F) Civatte bodies were also widely distributed across the epithelium, identified as round homogeneous and lightly stained globules. (Scale bar: 10 μm .) (G) Flattening and atrophy of the epithelial membrane with loss of rete ridge morphology that covered 75% to 100% of the biopsy specimen was considered the most severe. (Scale bar: 10 μm .)

histological changes (median score 1, G0). Histological changes prior to HCT were also mainly confined to G0-I ($n = 26$; median score 2, G0). The oral HCT control group was composed of patients clinically classified as no oral cGVHD and prior to oral cGVHD development. No significant differences were apparent between these 2 groups (data not shown), although biopsy specimens were classified from G0 to GIV ($n = 49$; median score 4, G1) (Figure 3A). Therefore, for all subsequent analyses, oral HCT control biopsy specimens were considered 1 group for comparison purposes. Patients who were clinically graded as possible oral cGVHD ($n = 44$; median score 7, GII) and defined oral cGVHD ($n = 78$; median score 10, GIII) showed significant variation across their histological scoring, with some defined oral cGVHD patients exhibiting few if any histological features, whereas others presented with many (Supplementary Figure S4B). None of the randomly selected biopsy specimens ($n = 46$) were found to be CMV-positive (data not shown). There was no statistically significant difference in histopathological scores between CMV-positive patients and CMV-negative patients within the large cohort or within their corresponding clinical group (Supplementary Figure S5).

Association of Oral Histopathology with cGVHD Activity and Histological Criteria

The NIH minimal histological oral GVHD criteria include lichenoid inflammation, exocytosis, and apoptosis [9]. In this study, specimens graded with lichenoid-like inflammation (score ≥ 2), focal to widespread exocytosis (score ≥ 1), and

variable apoptosis (score ≥ 1) were considered to meet these criteria. Within the oral HCT control group, 9 of 49 biopsy specimens (18.4% GII to GIV) met these criteria, and 15 of 44 (34.1% GII to GIV) in the possible oral cGVHD group were included. Approximately one-half of the defined oral cGVHD group (35 of 78; 44.9%, GII-GIV) met the NIH minimal histological criteria. No healthy specimens or specimens obtained prior to HCT identified as positive for the minimal features. Oral HCT controls, possible oral cGVHD, and defined oral cGVHD were assessed to establish whether any association existed between histopathological severity and overall cGVHD diagnosis (patients with no overall cGVHD score or those with overall cGVHD diagnosis). The only clinical classification exhibiting significantly different severity was oral HCT controls ($P = 0.001$) (Supplementary Figure S6). However, a significant association was apparent between oral histopathological severity and timing of biopsy with overall cGVHD activity (Figure 3B). Biopsy specimens obtained at cGVHD diagnosis and within the first year (median score 10; GIII) had a significantly higher histological score compared with those diagnosed with no overall cGVHD (median score 4; G1) ($P > 0.0001$).

Influence of Histological Features Correlated with Oral Clinical Status

To determine whether any specific histological feature contributed to oral cGVHD manifestations, each biopsy specimen was assessed with respect to its clinical classification against

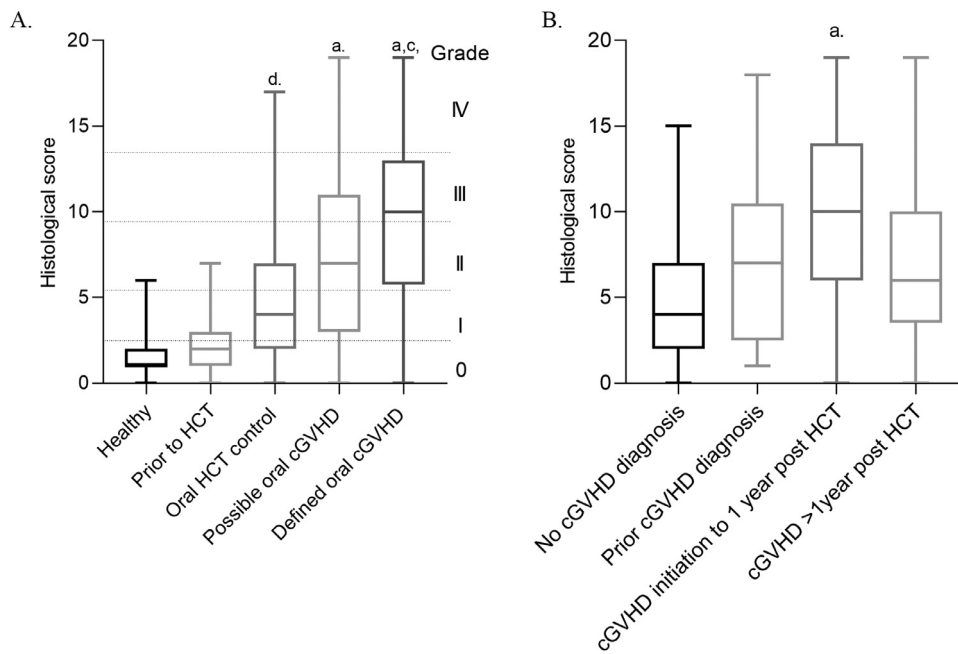


Figure 3. (A) Clinical assessment of the oral mucosa biopsy specimens ($n = 212$) plotted with respect to their corresponding histological score in the grading module (0 to 19, grade 0 to IV). Within the possible and defined oral cGVHD groups, a broad distribution of histological scores was observed, representing the dynamic nature of the active disease. Data are plotted presenting the median value and max to min. Comparisons: a = healthy versus possible, healthy versus defined, prior to HCT versus possible, prior to HCT; c = oral HCT control versus defined; and d = healthy versus oral HCT control, prior to HCT versus oral HCT control. (B) Oral histopathological severity score plotted against the overall cGVHD diagnosis and timing of biopsy ($n = 168$). Biopsy specimens obtained within the first year of diagnosis of overall cGVHD presented with a significantly higher histological score compared with those without any overall cGVHD. Data are plotted presenting the median value and range. Comparisons: a = no cGVHD diagnosis versus cGVHD initiation to 1-year post-HCT. a < 0.0001; b < 0.005; c < 0.001; d < 0.05.

the individual histological feature scores (Supplementary Figure S7). Inflammation, liquefaction degeneration, apoptosis, and basal membrane alterations were associated mainly with oral HCT controls, possible and defined, whereas exocytosis and flattening/atrophy were seen to a greater degree across all clinical presentations, including previous HCT and healthy. Analyzing the relationship between the histological grades (G0 to GIV) against the individual histological feature scores showed flattening/atrophy across all histological grades and possibly no significant influence on disease outcome/prediction (Supplementary Figure S8). The other features showed noteworthy correlation trends with respect to altered feature score and increased histopathological severity grades; however, no individual feature score was definitive of the final histological grade.

To further investigate and determine whether some features could have a higher predicted rate of occurrence, ROC plots were drawn for each feature according to the clinical presentation (Supplementary Figure S9). Data from biopsy specimens from prior to HCT and oral HCT controls served as control values. AUC values were calculated; all defined oral cGVHD values were $>.6$ and considered relatively accurate. LRs were calculated to assess any impact of histological feature due to transplantation alone (prior to HCT) or post-HCT (oral HCT controls). LRs for inflammation (13-fold), apoptosis (8.8-fold), atrophy (8.3-fold), and liquefaction degeneration (4.1-fold) all demonstrated high to moderate odds for developing these features in defined cGVHD when prior to HCT was used as a control. An LR for basal membrane alterations could not be determined, because specificity was 100%; however, when examining the impact of these features post-transplantation using oral HCT controls against defined oral cGVHD, basal membrane alterations had a 5-fold moderate LR, followed by inflammation at 3.8-fold and exocytosis and atrophy each at 3.1-fold, suggesting basal membrane

changes as a key feature that warrants further investigation to understand oral cGVHD severity.

DISCUSSION

Oral cGVHD is proposed to present as a trio of separate diseases, highlighting the importance of assessing the diagnostic and pathological features for each to support clinicians and researchers with patient alignment [10]. This study focused on oral mucosal cGVHD across an extensive cohort of allogeneic HCT recipients and controls to assess distinctive histopathological features, resulting in an objective tissue grading tool. The NIH Consensus Development Project has come a long way in terms of agreement on general terms for clinical diagnostics and routine care, as well as recommendations for clinical trials [5,6,26,27]. However, validated histopathological grading criteria in large cohorts for oral mucosal disease presentation is notable for its absence, leaving histopathological staging open for interpretation [9].

Herein we present a histological grading scheme that will be of value in the assessment of all HCT recipients, particularly those with normal mucosa and with nondiagnostic/active lesions but corresponding tissue changes, which will increase our knowledge of the complete biology following HCT and any possible chronic phase of GVHD, including refractory patients. This in turn will support comparisons between research findings and contribute to the emerging picture of oral GVHD, where the underlying mechanisms can be fully explored, leading to improved patient characterization to facilitate future investigations, treatments, and outcome prediction.

Application of the grading scheme across our full patient cohort ($n = 112$) and biopsy specimen collection ($n = 303$) has produced the most extensive oral mucosal cGVHD histopathological study to date. The assessment is strengthened by multiple sampling from some patients obtained before and after HCT, which provided valuable insight for evaluating the dynamic

changes in oral mucosal cGVHD progression over time, as well as the biological stages of GVHD. Recurring histological features were assessed, ranging from intraepithelial lymphocytes and band-like inflammatory infiltrate to atrophic epithelium with basal cell liquefaction degeneration including apoptosis, as well as separation of epithelium and formation of pseudo-rete ridges. Most features had been reported previously, but the pseudo-rete ridges were novel [28]. Additional features observed but not included in the scheme were leukoedema-like changes, as well as vacuolated signet ring-like cells. Some degree of epithelial edema has been reported by others but was deemed nonspecific pathology [14]. The appearance of vacuolated signet ring-like cells should not be mistaken for true basal cell degeneration, and thus they were considered artefacts [29,30]. We also noted pronounced signs of fibrosis in some specimens, which were discussed in terms of whether to include them as a hallmark for oral sclerotic GVHD. However, due to limited published data on oral GVHD submucosal fibrosis pathophysiology with mucosal scarring, and the difficulties involved in histological grading, these signs were not incorporated, but nonetheless, they do warrant further investigation [10,18].

NIH minimal histological criteria for active oral GVHD include lichenoid inflammation with lymphocytic exocytosis and apoptosis [9]. The magnitude of inflammatory clusters varied considerably across the cohort, as was expected owing to the dynamic nature of the disease and the time frame in relation to immunosuppressant medication [9,13,14,16,19,31]. Exocytosis and liquefaction degeneration also commonly occurred, consistent with previous findings [15,18]. Quantification of apoptosis, a key feature in GVHD pathology, has been proposed to finalize oral histological diagnoses [9,32]; however, considerable variation exists between published studies describing the extent of apoptosis within oral GVHD tissues [15–18]. Alterations to the basal membrane have not been discussed to the same degree, but reports of thickened and partial to complete clefting have been mentioned [15,17,19,33]. We identified 2 severity stages: an initial increased basal lamina thickness, followed by subsequent epithelial detachment with or without pseudo-rete ridges projecting into the lamina propria. In chronic oral inflammation, the basal membrane has been observed to be discontinuous and/or fragmented, with subsequent branching and pseudo-rete ridge formation [34,35]. Disruption might indicate biological stages of wound healing as observed with reticular OLP and leukoplakia lesions [35,36]. Basal membrane branching has been observed in both reticular and erosive OLP, whereas thickening was seen only in the presence of erosive disease [36].

Oral mucosal cGVHD histopathological guidelines include only minimal histological criteria for the active disease and are often used only as study criteria, with no reports of the final histopathological diagnosis of “no,” “possible,” and “likely” GVHD [17]. Our data show extensive variation of histological tissue reactions post-HCT in different clinical settings and over time, suggesting that these patients should not be stratified into one large oral cGVHD group. In the present investigation, one-third of the possible cGVHD group and almost one-half of the defined cGVHD group fell into GII to GIV and met the NIH minimal criteria [9]. However, a large proportion of biopsy specimens were outside the NIH remit [16,32,37]. Conversely, approximately one-fifth of oral HCT controls did meet the NIH criteria, in line with previous investigations into nonclinical GVHD affects on mucosa and skin [9,38,39]. Our data also support the proposal that histological changes may occur in the oral mucosa due to overall GVHD [37,40]. Oral HCT controls showed significantly altered histological changes when cGVHD

manifested concurrently in other organs. Histopathological score at time of cGVHD initiation and activation was significantly higher than that in non-cGVHD patients, regardless of oral status, with a score of \geq GIII considered “likely” cGVHD. Therefore, careful consideration of these criteria is required to ensure reliable patient stratification.

Unique to the present study was the availability of multiple biopsy specimens obtained from patients over time. The trends highlight the dynamic nature of the disease with individual patient response. However, owing to the study’s retrospective design, it was not possible to account for all parameters, including immune suppression time frame or subclassification diagnosis, which could have clarified the disease stages. Simultaneous confounding infections or tissue reactions could also affect interpretation of the histology [9]. CMV complications post-HCT range from asymptomatic viremia to end-organ enrollment and induced cytotoxic T cell response [41,42]. However, CMV infections within the oral mucosa probably are not as common, and none of our assessed tissue specimens were positive for CMV. No significant differences were observed in histopathological severity between CMV-positive and CMV-negative patients.

Further investigations are needed to validate and assess our grading scheme in the broader picture, including the correlation with oral clinical parameters as well as other organ scores with the significance of any confounding factors. However, of the histological features identified and within the literature, the likelihood of detecting liquefaction degeneration in specimens post-HCT was greater or equal to the likelihood of identifying exocytosis, suggesting that it also should be considered as a minimal feature similar to what is described by the NIH for skin [9]. NIH minimal skin cGVHD criteria contain more features than oral cGVHD, most of which are in line with our proposed histological grading scheme. The criteria for oral GVHD focuses on active disease, but no specific guidelines are suggested for prolonged phases, unlike for the skin [9].

In light of the current large cohort analyses, we propose a formalized histological grading system for oral mucosal cGVHD. Our grading tool can serve to describe the histopathological spectra in different clinical settings and be used to define minimal and specific criteria for oral mucosal cGVHD—G0 (no GVHD), GI (inconclusive/resolved/inactive GVHD), GII (possible GVHD), and GIII/GIV (likely GVHD)—to stratify patients into groups to support improved diagnostics and tailored treatments.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the clinical investigator, Elisabeth Boström, Department of Dental Medicine, Karolinska Institutet and Medical Dental Clinic Huddinge, Region Stockholm, for assistance with collecting healthy oral mucosal biopsy specimens.

Financial disclosure: Funding for this work was provided by Styrggruppen KI/Region Stockholm för Odontologisk Forskning (SOF), the Swedish Dental Society, and research grants from Karolinska Institutet.

Conflict of interest statement: There are no conflicts of interest to report.

Authorship statement: V.T. contributed to the study design, performed research, analyzed data, and wrote the manuscript. N.T. contributed to the study design, performed research, and reviewed the manuscript. G.W. contributed to the study design, performed research, and reviewed the manuscript. N. Y. performed research, analyzed data, and reviewed the manuscript. M.R. analyzed data and reviewed the manuscript. R.H.

performed research and reviewed the manuscript. K.G.L. performed research, analyzed data, and reviewed the manuscript. R.V.S. conceptualized the study, contributed to the study design, performed research, analyzed data, and wrote the manuscript.

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.bbmt.2020.06.031.

REFERENCES

- Gratwohl A, Baldomero H, Aljurf M, et al. Hematopoietic stem cell transplantation: a global perspective. *JAMA*. 2010;303:1617–1624.
- Arai S, Arora M, Wang T, et al. Increasing incidence of chronic graft-versus-host disease in allogeneic transplantation: a report from the Center for International Blood and Marrow Transplant Research. *Biol Blood Marrow Transplant*. 2015;21:266–274.
- Lee SJ. Classification systems for chronic graft-versus-host disease. *Blood*. 2017;129:30–37.
- Cooke KR, Luznik L, Sarantopoulos S, et al. The biology of chronic graft-versus-host disease: a task force report from the National Institutes of Health consensus development project on criteria for clinical trials in chronic graft-versus-host disease. *Biol Blood Marrow Transplant*. 2017;23:211–234.
- Filipovich AH, Weisdorf D, Pavletic S, et al. National Institutes of Health consensus development project on criteria for clinical trials in chronic graft-versus-host disease: I. Diagnosis and staging working group report. *Biol Blood Marrow Transplant*. 2005;11:945–956.
- Jagasia MH, Greinix HT, Arora M, et al. National Institutes of Health Consensus Development Project on Criteria for Clinical Trials in Chronic Graft-versus-Host Disease: I. The 2014 Diagnosis and Staging Working Group report. *Biol Blood Marrow Transplant*. 2015;21:389–401. e1.
- Lee SJ, Flowers MED. Recognizing and managing chronic graft-versus-host disease. *Hematology Am Soc Hematol Educ Program*. 2008:134–141.
- Treister N, Duncan C, Cutler C, Lehmann L. How we treat oral chronic graft-versus-host disease. *Blood*. 2012;120:3407–3418.
- Shulman HM, Cardona DM, Greenson JK, et al. NIH Consensus development project on criteria for clinical trials in chronic graft-versus-host disease: II. The 2014 Pathology Working Group report. *Biol Blood Marrow Transplant*. 2015;21:589–603.
- Bassim CW, Fassil H, Mays JW, et al. Oral disease profiles in chronic graft-versus-host disease. *J Den Res*. 2015;94:547–554.
- Shulman HM, Sullivan KM, Weiden PL, et al. Chronic graft-versus-host syndrome in man. A long-term clinicopathologic study of 20 Seattle patients. *Am J Med*. 1980;69:204–217.
- Sale GE, Shulman HM, Schubert MM, et al. Oral and ophthalmic pathology of graft versus host-disease in man: predictive value of the lip biopsy. *Hum Pathol*. 1981;12:1022–1030.
- Mattsson T, Sundqvist KG, Heimdahl A, Dahllöf G, Ljungman P, Ringdén O. A comparative immunological analysis of the oral mucosa in chronic graft-versus-host disease and oral lichen planus. *Arch Oral Biol*. 1992;37:539–547.
- Nakamura S, Hiroki A, Shinohara M, et al. Oral involvement in chronic graft-versus-host disease after allogeneic bone marrow transplantation. *Oral Surg Oral Pathol Oral Radiol Endod*. 1996;82:556–563.
- Soares AB, Faria PR, Magna LA, et al. Chronic GVHD in minor salivary glands and oral mucosa: histopathological and immunohistochemical evaluation of 25 patients. *J Oral Pathol Med*. 2005;34:368–373.
- Demarosi F, Lodi G, Carrassi A, Moneghini L, Sarina B, Sardella A. Clinical and histopathological features of the oral mucosa in allogeneic haematopoietic stem cell transplantation patients. *Exp Oncol*. 2007;29:304–308.
- Santos PS, Coracin FL, Barros JC, Gallottini MH. Histopathologic diagnosis of chronic graft-versus-host disease of the oral mucosa according to the National Institutes of Health consensus. *Einstein (Sao Paulo)*. 2014;12:204–210.
- Motta A, Zhan Q, Larson A, et al. Immunohistopathological characterization and the impact of topical immunomodulatory therapy in oral chronic graft-versus-host disease: a pilot study. *Oral Dis*. 2018;24:580–590.
- Horn TD, Rest EB, Mirenski Y, Corio RL, Zahurak ML, Vogelsang GB. The significance of oral mucosal and salivary gland pathology after allogeneic bone marrow transplantation. *Arch Dermatol*. 1995;131:964–965.
- Fujii H, Ohashi M, Nagura H. Immunohistochemical analysis of oral lichen-planus-like eruption in graft-versus-host disease after allogeneic bone marrow transplantation. *Am J Clin Pathol*. 1988;89:177–186.
- Shulman HM, Kleiner D, Lee SJ, et al. Histopathologic diagnosis of chronic graft-versus-host disease: National Institutes of Health Consensus Development Project on Criteria for Clinical Trials in Chronic Graft-versus-Host Disease: II. Pathology Working Group report. *Biol Blood Marrow Transplant*. 2006;12:31–47.
- Schubert MM, Williams BE, Lloid ME, Donaldson G, Chapko MK. Clinical assessment scale for the rating of oral mucosal changes associated with bone marrow transplantation. Development of an oral mucositis index. *Cancer*. 1992;69:2469–2477.
- Jenks GF. The data model concept in statistical mapping. *International Yearbook of Cartography*. 1967;7:186–190.
- de Oliveira Zakir JC, Casulari LA, Corrêa Rosa JW. Prognostic value of invasion, markers of proliferation, and classification of giant pituitary tumors, in a georeferred cohort in Brazil of 50 patients, with a long-term postoperative follow-up. *Int J Endocrinol*. 2016;2016: 7964523.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.
- Pavletic SZ, Lee SJ, Socie G, Vogelsang G. Chronic graft-versus-host disease: implications of the National Institutes of Health consensus development project on criteria for clinical trials. *Bone Marrow Transplant*. 2006;38:645–651.
- Lee SJ, Wolff D, Kitko C, et al. Measuring therapeutic response in chronic graft-versus-host disease. National Institutes of Health consensus development project on criteria for clinical trials in chronic graft-versus-host disease: IV. The 2014 Response Criteria Working Group report. *Biol Blood Marrow Transplant*. 2015;21:984–999.
- Fall-Dickson JM, Pavletic SZ, Mays JW, Schubert MM. Oral complications of chronic graft-versus-host disease. *J Natl Cancer Inst Monogr*. 2019;2019. lgz007.
- Garijo MF, Val D, Val-Bernal JF. Pagetoid dyskeratosis of the nipple epidermis: an incidental finding mimicking Paget's disease of the nipple. *APMIS*. 2008;116:139–146.
- Petrilli G, Fisogni S, Rosai J, et al. Nuclear bubbles (nuclear pseudo-pseudoinclusions): a pitfall in the interpretation of microscopic sections from the thyroid and other human organs. *Am J Surg Pathol*. 2017;41:140–141.
- Nicolatou-Galitis O, Kitra V, Van Vliet-Constantinidou C, et al. The oral manifestations of chronic graft-versus-host disease (cGVHD) in paediatric allogeneic bone marrow transplant recipients. *J Oral Pathol Med*. 2001;30:148–153.
- Nakhleh RE, Miller W, Snover DC. Significance of mucosal vs salivary gland changes in lip biopsies in the diagnosis of chronic graft-vs-host disease. *Arch Pathol Lab Med*. 1989;113:932–934.
- Barrett AP, Bilous AM. Oral patterns of acute and chronic graft-v-host disease. *Arch Dermatol*. 1984;120:1461–1465.
- Shklar G, Flynn E, Szabo G. Basement membrane alterations in oral lichen planus. *J Invest Dermatol*. 1978;70:45–50.
- Larjava H, Haapasalmi K, Salo T, Wiebe C, Uitto VJ. Keratinocyte integrins in wound healing and chronic inflammation of the human periodontium. *Oral Dis*. 1996;2:77–86.
- Jungell P, Malmström M, Wartiovaara J, Konttinen Y, Sane J. Ultrastructure of oral leukoplakia and lichen planus. I. Basal region and inflammatory cells. *J Oral Pathol*. 1987;16:170–178.
- Prochorec-Sobieszek M, Nasilowska-Adamska B, Szumera-Cieckiewicz A, et al. The significance of oral labial biopsy in hepatic graft-versus-host disease diagnosis in patients following allogeneic hematopoietic stem cell transplantation—a preliminary report. *Ann Transplant*. 2012;17:85–92.
- Heimdahl A, Johnson G, Danielsson KH, Lönnqvist B, Sundelin P, Ringden O. Oral condition of patients with leukemia and severe aplastic anemia. Follow-up 1 year after bone marrow transplantation. *Oral Surg Oral Med Oral Pathol*. 1985;60:498–504.
- Vassallo C, Brazzelli V, Alessandrino PE, et al. Normal-looking skin in onco-haematological patients after allogeneic bone marrow transplantation is not normal. *Br J Dermatol*. 2004;151:579–586.
- Gonçalves Resende R, de Fátima Correia-Silva J, Arão TC, et al. Oral cGVHD screening tests in the diagnosis of systemic chronic graft-versus-host disease. *Clin Oral Investig*. 2012;16:565–570.
- Ljungman P, Hakki M, Boeckh M. Cytomegalovirus in hematopoietic stem cell transplant recipients. *Hematol Oncol Clin North Am*. 2011;25:151–169.
- Cho SY, Lee DG, Kim HJ. Cytomegalovirus infections after hematopoietic stem cell transplantation: current status and future immunotherapy. *Int J Mol Sci*. 2019;20:2666.