



UNIVERSITÀ
DI SIENA
1240

University of Siena – Department of Medical Biotechnologies
Doctorate in Genetics, Oncology and Clinical Medicine (GenOMeC)
XXVIII cycle (2012-2015)

Scientific disciplinary sector: MED/16 Reumatology

**Identification of HLA functional motifs associated with severe
systemic sclerosis and sclerotic-graft versus host disease:
The shared epitopes of fibrosis**

Tutor

Prof. Marco Matucci Cerinic

PhD Candidate

Jelena Blagojevic

Coordinator: Prof. Alessandra Renieri

Academic Year 2015/2016



27 rue du Faubourg Saint-Jacques
75679 PARIS CEDEX 14
Standard : 01 58 41 41 41

Hôpital COCHIN
Pavillon HARDY
Service de Rhumatologie A
Chef de Service :
Pr A. KAHAN
01 58 41 25 50

Adjoint :
Pr Yannick ALLANORE
01 58 41 25 72
RPPS : 10001476059
N° Ordre des médecins :
Paris 75/66494

Institut Cochin, INSERM U1016
*Genetics of the immune system and
chronic rheumatic conditions*
Pavillon Roussy
4ème étage
Tél: 01 40 51 65 34
Fax: 01 40 51 66 41

Secrétariat
Mme Carole DESBAS
Tél. : 01 58 41 25 63
Fax : 01 58 41 26 24
carole.desbas@cch.aphp.fr

To whom it may concern,

Systemic sclerosis (SSc) is characterized by early vascular damage and immune dysregulation followed by tissue fibrosis affecting many organs, in particular, the skin, lungs, heart, kidneys and digestive tract. The epidemiological data are consistent to define SSc as an orphan disease. The combination of environmental and stochastic factors converging upon individuals of a particular genetic background seems to contribute to the development of this disease. Subtle combinations and interactions may also contribute to the heterogeneity of the disease that is a hallmark of this condition.

The association between HLA molecules and SSc was established in the 1970s. SSc is associated with class II HLA molecules for which strong genotypic variation has been found between ethnic groups and, in some cases with gender (DQA1*05:01 is more strongly associated with the disease in men than in women; Lambert et al., 2000). The genotype also varies particularly strongly as a function of the presence of different types of autoantibodies associated with SSc: anti-topoisomerase I (topo) with DRB1*11*15:02 (Kuwana et al., 1999), DQB1*03*04 (Ueki et al., 2001), DPB1*13:01 and DPB1**09:01 (Gilchrist et al., 2001) and anti-centromere antibodies (ACA) with DRB1*01:01, DRB1*01:04, DRB1*01:08, DQB1*05:01, DPB1*04:02 (Kuwana et al., 1999; Gilchrist et al., 2001). A large study of HLA class II genes (DRB1, DQB1, DQA1, DPB1) was recently carried out in 1300 SSc cases (961 Caucasians) and 1000 controls (Arnett et al., 2010). The strongest associations identified were those with the DRB1*11:04, DQA1*05:01 and DQB1*03:01 haplotypes and the DQB1 allele. However, many other alleles or haplotypes have also been associated with certain subtypes of the disease, particularly those defined on the basis of the presence of autoantibodies. In a similar study, 944 Caucasian SSc patients and 1320 unaffected controls from Italy and Spain were examined. The association of DRB1*11:04, DQA1*05:01 and DQB1*03:01 haplotype with SSc was confirmed in this study (Beretta et al. 2012).

In the first GWAS (Zhou et al., 2009), the strongest association ($P = 8.16 \times 10^{-13}$) identified was that with five SNPs in the HLA-DPB1 and -DPB2 regions. A more detailed study of this region subsequently identified rs3128930, rs7763822 and rs7764491 as responsible for this association, particularly in cases of anti-topoisomerase I antibody detection. In the American replication cohort, rs7763822 and rs7764491 were the most strongly associated with SSc. These findings confirm the relevance of HLA-DPB1 for SSc and especially topo-positivity. In the two additional, large-scale GWAS (Radstake et al., 2010 and Allanore et al. 2011), the strongest association was also that with the class II HLA region. These findings also reflect the considerable differences in HLA genetic background between ethnic groups and the tremendous

heterogeneity of the disease itself. Finally, the complex genetic structure of the HLA system in terms of LD complicates fine analyses and the identification of causal variants

The candidate has written a precise and easy to read introduction about SSc, its genetic bases and the HLA structure together with its implication in various autoimmune diseases. She has particularly introduced HLA genomic bases, and its key role in a related disease that is rheumatoid arthritis.

Chronic graft versus host disease (cGVHD) is one of the most severe complications of the allogeneic hematopoietic stem cell transplantation (HSCT) and is currently the main cause of long-term morbidity and mortality following HSCT. Its clinical characteristics resemble autoimmune and other immunologic disorders such as SSc. Sclerotic (sclerodermatous) GVHD (sclGVHD) is a distinctive clinical form of cGVHD with sclerotic features in the skin characterized by thickened, tight, and fragile skin. Similarities and differences between SSc and sclGVHD are well discussed in the manuscript.

The hypothesis of this project is to use sclGVHD and the diffuse cutaneous as similar severe fibrotic skin diseases and to compare combined patients to SSc patients with milder fibrotic subset. One may ask why patients with non fibrotic form of GVHD were not considered as controls and also why healthy individuals were not included. Therefore, the aim of this study was to identify the HLA class II functional motifs (shared epitopes) associated with severe skin fibrosis in SSc and scl-GVHD.

Pr Del Galdo previously used this strategy to link SSc with another fibrotic disease that is nephrogenic fibrosis. He also worked intensively on murine GVH. Pr Matucci Cerinic is a recognized international expert working both on pathogenesis and clinical expression.

The candidate was co-supervised, and her exact role should be further explained to know if she contributed to the all the steps, including recruitment and phenotyping of the patient, molecular biology experiments, statistical analyses.

Based on the hypothesis that the analysis of the HLA hypervariable regions in SSc and sclGVHD patients may reveal the existence of common functional motifs (shared epitopes), similar to those demonstrated in RA, the authors used the following strategies:

- high resolution four-digit genotyping in severe SSc and sclGVHD patients
- in-depth analyzes of amino acidic sequences corresponding to the hypervariable regions of β chains of HLA class II molecules (DPB1, DQB1 and DRB1)
- quantitative analysis of the similarity of hydrogen bonds strength within the different populations

The results section starts with the phenotypic data of 45 patients with sclGVHD from the US, Bethesda center, for whom detailed examination

regarding skin fibrotic involvement is provided. It seems that 80 SSc patients were recruited in Milano (Italy) but that only 70 could be included; this should be clarified (pages 55-57). Detailed phenotypic characteristics are provided. They show discrepancies between SSc of diffuse subtype and sclGVH for sex and ethnicity. High resolution HLA typing of SSc patients and of sclGVHD patients without 4-digit HLA typing was performed by commercially available kits (Fujirebio Diagnostics):

The amino acids corresponding to the hypervariable regions were located in the different positions:

- HLA-DPB1: 8-11, 35-36, 55-57, 65-69 and 84-88
- HLA-DQB1: 26-37, 52-57, 70-74
- HLA-DRB1: 9-13, 25-28, 67-74

Despite the strategical decision to analyse SclGVHD and DcSSc patients together as they some table still show separately sclGVH, DcSSc and LcSSc. Then the various results are presented according to the method and type of analyses.

Comparison of frequencies between the groups

According to various statistical analyses and level of correction, various associations are found. The strengths and limitations of the various methods could be introduced.

HLA - DRB1 position 71 and HLA - DPB1 position 11 show different distributions of the amino acid frequencies but using a less stringent method differences of amino-acidic frequencies between groups at HLA-DQB1 position 26 and HLA - DRB1 position 67 were highlighted. This could be further commented according to the above comment about the statistics.

Then, whether there was any combination of amino acids associated with risk of developing severe fibrotic phenotype characterized by severe immune-mediated fibrosis is queried.

Association of pairs of amino-acids with risk of developing fibrotic phenotype

For HLA-*DPB1* a DEDEAV amino acid sequence was observed, 37 of these 56 patients (66%) belonged to the group 2 (had non-fibrotic phenotype) while only 18 of the other 51 patients who did not present this DEDEAV amino acid sequence (35%) were in the group 2.

Regarding the HLA - DQB1 analysis, Amino acid G at position 26, produced a difference: it was present in 51 patients of whom 35 (69%) had non-fibrotic phenotype (group 2), compared to 20 out of 53 patients (38%) with other amino acids at position 26 who presented non-fibrotic phenotype ($p=0.0018$).

Regarding HLA - DRB1, amino acid F at position 67 was present in 62 patients; 41 of them (66%) had non-fibrotic phenotype (group 1) versus 14 out of 43 (33%) of patients with other amino acids at the same position ($p=0.0008$).

Association of triplets of amino-acids with risk of developing fibrotic phenotype

The strongest association for HLA-DPB1 was observed between two triplets having L, A and A at positions 11, 36 and either 55 or 56 and

non-fibrotic phenotype. The second strongest DPB1 triplet association with the non-fibrotic phenotype (group 2) was shared by a set of 21 triplets that were all sub-sequences of a combination of 10 specific amino acids across ten positions (VFVDEIDEAV at positions 8,35,36,55,56,65,84-87).

The strongest association for DQB1 was shared by 36 triplets which were present in 35 out of the 55 patients (64%) with non-fibrotic phenotype versus 16 out of the 49 patients (33%) with fibrotic phenotype (group 1). Further examination of the triplets revealed that they were all sub-sequences of a combination across eight positions, 26, 28, 37, 52, 55, 56, 74 and 75, containing amino acids G, T, Y, P, R, P, S and V.

The most significant association for DRB1 with the study groups consisted in the association of non-fibrotic phenotype (group 2) with five triplets which had amino acids F and D at positions 67 and 70 ($p < 0.00042$).

Qualitative analysis of amino-acids

DPB1 position 11 was the only one with a significant difference between two study groups in the family frequencies ($p = 0.033$). When correction was applied, differences in the family frequencies for the HLA - DQB1 position 26 resulted also significant. The classifications of hydrophobic versus non-hydrophobic and polar versus non-polar did not produce differences that reached the above level of statistical significance.

Altogether, the results show some common amino acidic sequences within the hypervariable regions of β chains of HLA – DP, HLA – DQ and HLA – DR molecules to be present in lcSSc patients with minimal skin fibrosis but not in SSc and sclGVHD patients with severe skin fibrosis. Nevertheless, some limitations must be raised including that the working hypothesis of merging sclGVH and DcSSc can be challenged because of various clinical expression and tissular damages. Nevertheless, it is original and could get insight into severe fibrotic disease but in that sense, some even extreme phenotype could be isolated within sclGVH+dcSSc patients (highest score or correlation with fibrotic skin damages using eQTL). The control group can also be challenged as above stated. Finally, the methodology contains some biases like ethnicity, but the biggest limitation is probably the relative small sample size for such heterogeneous conditions. Discrepancies with previous studies would deserve more discussions in the ms. The part about the role of autoantibodies might be tempered because in SSc, several studies particularly about anti-topoisomerase suggest a biological role of these autoantibodies. Some more details about the statistical analyses could be added to further give priorities to some results according to the relevance of the analyses.

The candidate Jelena Blagojevic has published so far 4 studies referenced in Pubmed about systemic sclerosis which shows her interest in this disease. The publication plan of the herein work is not introduced but a Review paper could be done with the nice and complete introduction of the present ms. The publication of the original

work may be more challenging and one might argue that enlarging the cohort of patients and probably also of the controls might be needed. As a final comment about this nice and interesting work, the candidate should give more insight about her own contribution. I support that Mrs Blagojevic can defend orally her thesis according to the content of the present document.

Paris 6th of June 2016
Pr Yannick Allanore

G.H. COCHIN ST-VINCENT DE PAUL

Service de Rhumatologie A

Professeur Y. ALLANORE

PU-PH

27, rue du Faubourg Saint Jacques

75679 PARIS CEDEX 14

Sec.: Tél. 01 58 41 25 63 : e-mail : carole.desbus@coch.saphp.fr



Department of Experimental and Clinical Medicine
Division of Rheumatology

Florence, June 7th, 2016.

To:

Prof Yannick Allanore

University Paris Descartes

University Hospital Cochin, Rheumatology Service A

Dear prof Allanore,

Thank you for your comprehensive, detailed and accurate revision of my PhD thesis.

Please find enclosed the responses on your comments and questions:

- 1. The candidate has written a precise and easy to read introduction about SSc, its genetic bases and the HLA structure together with its implication in various autoimmune diseases. She has particularly introduced HLA genomic bases, and its key role in a related disease that is rheumatoid arthritis.*
- 2. Similarities and differences between SSc and sclGVHD are well discussed in the manuscript.*
- 3. The hypothesis of this project is to use sclGVHD and the diffuse cutaneous as similar severe fibrotic skin diseases and to compare combined patients to SSc patients with milder fibrotic subset. One may ask why patients with non fibrotic*

form of GVHD were not considered as controls and also why healthy individuals were not included.

Response: This is a fair concern from the reviewer. The leading hypothesis was to identify common HLA motifs between fibrotic conditions with similar severity but different etiology. This was to determine whether some pilot data could be achieved to pursue funding for a larger study. In this view, and given the known overall heterogeneity of the hypervariable regions within the HLA we restricted the analysis to a discovery cohort of only sclGVHD. Further, these were the only available for four-digit HLA typing at NCI in Bethesda.

4. *The candidate was co-supervised, and her exact role should be further explained to know if she contributed to the all the steps, including recruitment and phenotyping of the patient, molecular biology experiments, statistical analyses.....As a final comment about this nice and interesting work, the candidate should give more insight about her own contribution.*

Response: This study was initially created as a collaborative project between the University of Leeds and National Cancer Institute in Bethesda, USA with an aim to identify shared amino acid sequences in the hypervariable regions of HLA class II molecules between the sclGVHD patients and SSc patients who received autologous haematological stem cell transplantation (HSCT) because of severe skin fibrosis. The beginning of the project was temporary suspended due to the difficulty in recruiting of SSc patients who performed HSCT and had DNA available for the HLA analysis.

I reviewed the initial protocol and suggested to modify it by including a SSc group comprising both patients with severe skin involvement as surrogate of the original group and a control group with limited SSc in order to obtain more informative results. I identified dr Lorenzo Beretta in Milan who recently

published interesting scientific papers on HLA class II genotyping in SSc and contacted him together with the project coordinator inviting him to participate in the study. Further, I have suggested to modify the qualitative analysis of the amino acid families by subgrouping them in more coherent classes, and I have coded the sequence for the quantitative analysis.

SclGVHD patients were recruited from the pre-existing study on natural history of chronic GVHD of NCI (study 04-C-0281) and already had extensive clinical, laboratory, instrumental and histopathological assessment as for study protocol, as described in the Materials and Methods section (page 49, rows 12-25). I identified patients with sclGVHD and skin involvement greater than 50% body surface area to be included into the present study.

The genotyping of all sclGVHD patients had been performed as a routine procedure prior to HSCT and for many without the high resolution required for the quantitative analysis. I revised the results of the available genotyping, translated the genotyping in aminoacidic sequences and tabled the hypervariable region of interest in an easy to access data format. I then flagged to the NCI dedicated technician (Sharon Adams) all DNA that needed re-genotyping in high resolution because of ambiguity in the aminoacidic sequence. SSc patients were recruited by Dr Beretta. I reviewed their clinical, laboratory and instrumental data and divided patients in limited and diffuse subset according to their clinical characteristics. I further divided dSSc patients into severe and not severe fibrotic phenotype according to the extent of skin involvement. The genotyping of SSc patients was performed by the team of Dr Beretta.

I performed HLA data alignment and identification of amino acids belonging to the hypervariable regions of HLA class II molecules. I ideated and performed preliminary data analysis. I also performed the preliminary statistical analysis.

Further statistical analysis has been done with the collaboration of Dr David Venzon, statistician of NIH in Bethesda, USA. I personally created all the queries and the code for quantitative analysis to be run by the statistician. Dr Venzon

suggested the types of statistical tests to be used and helped in performing the statistical analysis.

5. *It seems that 80 SSc patients were recruited in Milano (Italy) but that only 70 could be included; this should be clarified.*

Response: Eighty SSc patients were enrolled. On analysis of their HLA data I noted that ten of them had ambiguous alleles and unclear genotyping and were excluded from the study. I clarified this in the Materials and Methods section (page 50, rows 19-20 and page 51, row 1).

6. *Despite the strategical decision to analyse SclGVHD and DcSSc patients together as they some table still show separately sclGVH, DcSSc and LcSSc.*

Response: We thank the reviewer for noticing this putative inconsistency. This choice was done only for two tables with the descriptive results of the HLA genotyping in two diseases (sclGVHD and SSc) in order to make it more clear for the further comparison with the already published results on genotyping.

7. *Comparison of frequencies between the groups. According to various statistical analyses and level of correction, various associations are found. The strengths and limitations of the various methods could be introduced.....Some more details about the statistical analyses could be added to further give priorities to some results according to the relevance of the analyses.*

Response: Following the reviewer suggestion more details about the statistical analysis and various test have been added in the Statistical analysis section (page

57, rows 27-29 and page 58, rows 1-6). Specifically: “The multiple comparison corrections were performed with Hochberg correction and False Discovery Rate (Benjamini-Hochberg) correction. Hochberg correction, a step-up version of Bonferroni’s test was performed to control type I error rate and to reduce the risk of false positives. However, since this method may lead in turn to high number of false negatives for a large number of comparisons, we also performed alternative multiple comparison correction with less sensitive, Benjamini-Hochberg or False Discovery Rate (FDR) correction. FDR controls the expected proportion of "discoveries" (significant results) that are in reality false positives”.

8. *Discrepancies with previous studies would deserve more discussions in the ms. The part about the role of autoantibodies might be tempered because in SSc, several studies particularly about anti-topoisomerase suggest a biological role of these autoantibodies.*

Response: We did not think of this thesis work as a manuscript style. Nevertheless, following the suggestion, I have included in the discussion the following paragraphs on pages 81-84.

“They identified two common amino acidic motifs codified by these alleles: FLEDR at positions 67 – 71 of the HLA-DR β chain and TRAE LDT at positions 71 – 77 of the HLA-DQ β chain (267). When the patients were divided by clinical subset, both sequences were associated only with diffuse SSc. When the autoantibody status was considered, the association was confirmed only in ATA positive dSSc patients.” (page 80, rows 14-19).

“It is important to note that our study population was different from the population studied by the French authors. First of all, in the study performed by Azzouz et al, the association of the shared epitopes with dSSc was only present

in ATA positive patients, suggesting that FLEDR and TRAE LDT motifs might be associated with ATA production, rather than with clinical subset and with the extension of skin fibrosis. In the present study we analysed the clinical outcome (severe skin fibrosis) in patients with two different fibrotic diseases. Only in SSc the extent of skin fibrosis is correlated with ATA production whilst autoantibodies have not been demonstrated to play a role in the pathogenesis of sclGVHD (171). It is clear that potential common pathogenic mechanism underlying fibrosis in SSc and sclGVHD is not related to ATA production, and this may completely explain the apparent discrepancy of our results with those reported by Azzouz. To note that in our study population 80% of dSSc patients were ATA positive and none of sclGVHD patient were ATA positive “. (page 80, rows 23-28 and page 81, rows 1-7).

“Mean values of mRSS have not been reported in the paper published by Azzouz, therefore it is not possible to compare the severity of skin fibrosis between our study population and those described by the French authors”. (page 81, rows 11-13).

“We have not performed the subanalysis in the limited SSc ACA positive patients, but in our study population more than 80% of ISSc patients were ACA positive”. (page 81, rows 18-20).

“Differently from Mayes et al, we observed the protective effect of amino acid F at position 67 against dSSc and sclGVHD. As previously mentioned, our study population comprised also patients with sclGVHD in which ATA do not play a pathogenic role. Amino acids at position 13 were not distributed differently between the two groups in our study. To note that Mayes et al compared SSc patients to healthy subjects and the methodology for HLA typing was different between the two studies (ImmunoChip versus high resolution HLA typing)” (page 82, rows 3-9).

“For this reason we did not perform comparative analysis between patients and healthy subjects”. Importantly, the shared epitopes seem to play a role in ATA production in SSc, as previously mentioned.” (page 83, rows 1-4).

“In SSc, the presence of specific autoantibodies is strongly associated with distinct clinical phenotypes, but there is no unanimous consensus on their direct role in the pathogenesis of SSc or its clinical manifestations (**288**). Nevertheless, Fava et al have recently shown that the frequency of circulation DNA topoisomerase I specific CD4 T cells predicts presence and progression of interstitial lung disease in SSc (**289**). In addition, it has been reported that T cell lines from SSc patients and healthy controls recognize different epitopes on DNA topoisomerase I (**290**). It might be hypothesized that shared epitopes on HLA molecules bind specific portions of DNA topoisomerase and present them to T cells. To note that not all SSc patients and not all patients with diffuse SSc patient are ATA positive, and that other mechanism might be involved in pathogenesis of the disease and of its fibrotic manifestations.” (page 83, rows 5-15).

Finally, two new references (289 and 290) have been added: (page 119, rows 23-28).

Kind regards

Jelena Blagojevic

Dr Francesco Del Galdo
Lead, Scleroderma Research Program
Leeds Institute of Rheumatic and Musculoskeletal Medicine

6th Floor, Welcome Trust Brenner Building
University of Leeds
Becket Street
LS97TF

F.delgaldo@leeds.ac.uk

Work: 01133438412

2th June 2016

RE: Dr Jelena Blagojevic, PhD Thesis:

"Identification of HLA functional motifs associated with severe systemic sclerosis and sclerotic graft versus host disease: The shared epitopes of fibrosis."

Dear Sir/Madam

I have read and examined the PhD thesis from Dr Blagojevic on her study to identify putative shared functional motifs within the HLA of patients with similar level of severe fibrosis although from different conditions.

The Thesis is well structured with comprehensive introduction and background to put the reader in context. The overall style is clear and straight to the point and the references are chosen carefully and accurately.

The work itself is an extremely original and complex approach to tackle an unresolved step in the pathogenesis of immune driven fibrosis.

Although the genetic techniques used are relatively common and simple, relying on the standard of HLA genotyping, the statistical analysis and the angle of interpretation of the findings are extremely interesting.

In particular the identification of motifs that may carry a protective role for tissue fibrosis is quite a strong finding that will be surely published on good impact journals and it may have long lasting repercussion on the way we diagnose and stratify for risk of severity patients with either diffuse systemic Sclerosis or with Chronic Allograft rejection.

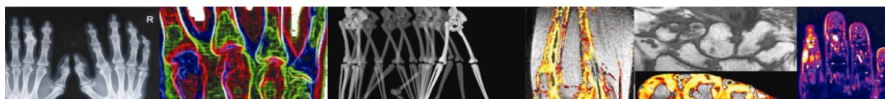
The discussion is well focused as well and also humble in pre-empting the weaknesses of the work.

I will be delighted to be involved in the thesis defense and go in detail of the study performed.

Shall you need any further information please do not hesitate to contact me

Yours Sincerely

Francesco Del Galdo



Leeds Musculoskeletal
Biomedical Research
Unit



Department of Experimental and Clinical Medicine
Division of Reumatology

Florence, June 3th, 2016.

To:

Prof Francesco Del Galdo

Leeds Institute of Rheumatic and Musculoskeletal Medicine

University of Leeds

Dear prof Del Galdo,

Thank you for reviewing my thesis. I have really appreciated your comments on my work.

Kind regards

Jelena Blagojevic

Table of contents

1.	INTRODUCTION.....	3
1.1.	Human leukocyte antigen structure and function and shared epitope hypothesis	3
1.1.1.	<i>Human leukocyte antigen class II genes and their products</i>	<i>3</i>
1.1.2.	<i>Antigen presentation: interaction between HLA molecules, peptides and T cell receptors ..</i>	<i>5</i>
1.1.3.	<i>HLA class II diversity</i>	<i>10</i>
1.1.4.	<i>Shared epitope hypothesis</i>	<i>13</i>
1.2.	Systemic sclerosis and chronic sclerotic graft versus host disease	18
1.2.1.	<i>Systemic sclerosis</i>	<i>18</i>
1.2.2.	<i>Pathogenesis of fibrosis in SSc.....</i>	<i>20</i>
1.2.3.	<i>Sclerotic graft versus host-disease</i>	<i>26</i>
1.2.4.	<i>Pathogenesis of cGVHD and sclGVHD</i>	<i>27</i>
1.2.5.	<i>SSc versus sclGVHD: similarities and differences.....</i>	<i>34</i>
1.3.	HLA class II alleles in systemic sclerosis and sclGVHD.....	37
1.3.1.	<i>HLA class II alleles in SSc.....</i>	<i>37</i>
1.3.2.	<i>SSc and HLA-DRB1 alleles</i>	<i>37</i>
1.3.3.	<i>SSc and HLA-DPB1 alleles</i>	<i>41</i>
1.3.4.	<i>SSc and HLA-DQB1 alleles.....</i>	<i>42</i>
1.3.5.	<i>SSc and HLA-DQA1 alleles</i>	<i>45</i>
1.3.6.	<i>SclGVHD and HLA class II alleles.....</i>	<i>45</i>
2.	AIM	46
3.	MATERIALS AND METHODS	49
3.1.	Patients.....	49
3.1.1.	<i>SclGVHD patients selection and evaluation.....</i>	<i>49</i>
3.1.2.	<i>SSc patients selection and evaluation</i>	<i>50</i>
3.2.	HLA typing	51
3.3.	HLA data alignment and analysis	52
3.3.1.	<i>Qualitative analysis</i>	<i>53</i>
3.3.2.	<i>Quantitative analysis.....</i>	<i>55</i>
3.3.3.	<i>Statistical analysis</i>	<i>57</i>
4.	RESULTS	59
4.1.	Comparison of amino-acidic frequencies between the groups	65
4.2.	Association of pairs of amino-acids with risk of developing fibrotic phenotype	66
4.3.	Association of triplets of amino-acids with risk of developing fibrotic phenotype	68
4.4.	Qualitative and quantitative analysis of amino acids associated with fibrotic/non fibrotic phenotype	74
4.4.1.	<i>Qualitative analysis of amino-acids.....</i>	<i>74</i>
4.4.2.	<i>Quantitative analysis of amino-acids</i>	<i>76</i>
5.	DISCUSSION	80
6.	CONCLUSIONS.....	87
7.	REFERENCES	88

1. INTRODUCTION

1.1. Human leukocyte antigen structure and function and shared epitope hypothesis

1.1.1. Human leukocyte antigen class II genes and their products

The human leukocyte antigens (HLAs) are cell membrane glycoproteins with a central role in the cell-mediated adaptive immune response, which main role is to present self and foreign antigens in the form of peptides to T lymphocytes (1). They are the human version of the major histocompatibility complex (MHC), the component of immune system that exists in most vertebrate species. HLAs were originally defined as the most important molecules involved in the rejection of transplanted tissues, but were later implicated in a number of human pathologies.

HLA gene complex is a family containing multiple genes located at the short arm of chromosome 6 in the distal portion of the 21.3 band. It consists of approximately 4 mega bases (Mb) and contains around 260 genes with variable expression and function, arranged to form three regions due to their different loci (class I, II and III regions) (Figure 1.1) (1, 2).

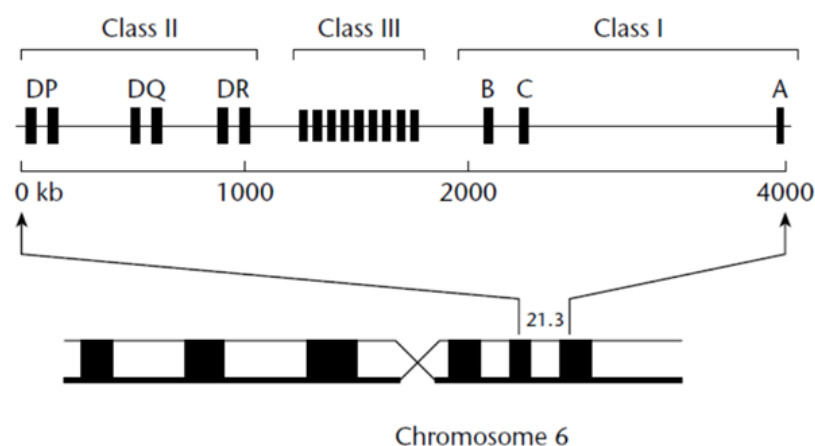


Figure 1.1 - HLA gene complex (adapted from reference 1)

Most of these genes are inherited jointly as a haplotype. Even though it occupies only 0.13% of the human genome, HLA complex is associated with more than 100 diseases, mostly autoimmune, and with drugs hypersensitivities (3, 4).

HLA genes are transmitted as codominant traits, thus an individual can express two alleles at each locus derived inherited from maternal and paternal haplotype.

The class II region contains classical class II alpha and beta chain genes (HLA-DP, -DQ and -DR) and non-classical HLA class genes (HLA-DM and -DO), which have very limited polymorphism (5). Apparently all genes in the class II region have functions related to antigen processing or presentation, except for bromodomain containing 2 (BRD2) gene which encodes for a transcriptional factor with widespread specificity involved in remodelling of chromatin complexes (6).

The class II molecules are heterodimers; consisting of non-covalently associated α and β chains encoded by the HLA class II classical genes (HLA-DRA, DQA, DPA and HLA-DRB, DQB and DPB, respectively) (5). The α chain has higher molecular weight (around 34 kDa) compared to β chain (approximately 29 kDa) (2). Both chains have extracellular, transmembrane and intracytoplasmic region and consist of four domains: the peptide-binding domain ($\alpha 1$ or $\beta 1$), the immunoglobulin-like domain ($\alpha 2$ or $\beta 2$), the transmembrane domain and intracytoplasmic tail (Figure 1.2) (1, 2).

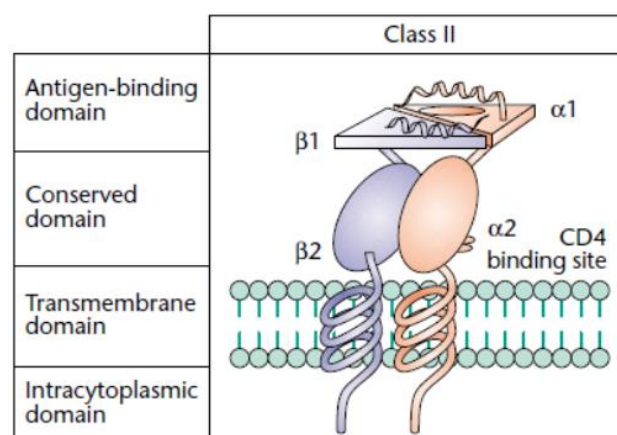


Figure 1.2 - HLA class II molecule structure and three-dimensional folding (adapted from reference 1)

Class II molecules are constitutively expressed on selected cells with immune function, such as B lymphocytes, and other “professional” antigen-presenting cells such as

monocytes, macrophages and dendritic cells, and activated T lymphocytes. However, under some circumstances their expression may be transiently induced on most cells by cytokines, such as tumour-necrosis factor α or interferon γ (7).

1.1.2. Antigen presentation: interaction between HLA molecules, peptides and T cell receptors

HLA molecules have a pivotal role in the presenting of the peptide antigens to T cells, a process that initiate the cell-mediated arm of the adaptive immune response. The T cells are able to recognize only peptide antigens that are displayed on the membranes of our cells. These peptides derive from the partial intracellular proteolysis and can origin from pathogens which replicate within cells, or from pathogens or their products that cells have internalized from the extracellular environment (8).

In the class II pathway, most antigen peptides derive from the exogenous proteins, internalized from the external environment by endocytosis or phagocytosis and then degraded by lysosomes and phagolysosomes respectively (Figure 1.3) (9, 10).

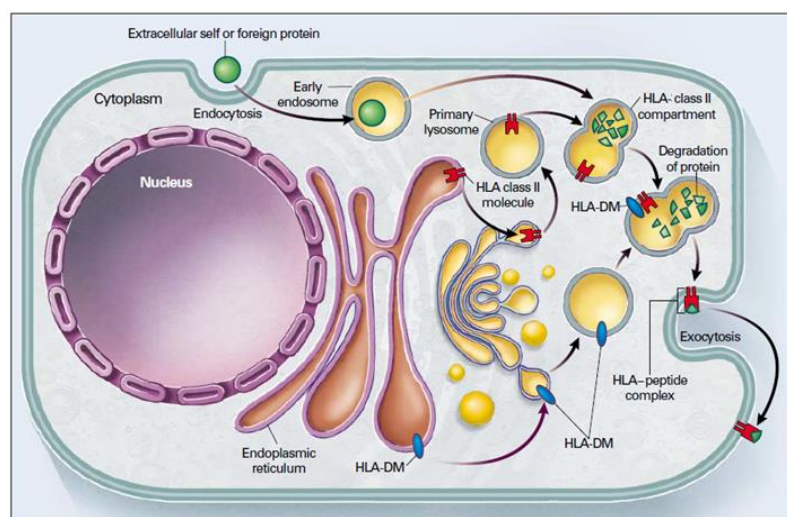


Figure 1.3 - Processing of extracellular protein in HLA class II pathway. Extracellular protein phagocytosis and degradation, peptide binding to HLA class II molecule and exposure on the cell surface (adapted from reference 10).

Class II molecules are associated with a specific chaperone, the invariant chain (I chain), which promotes their folding, prevents premature loading of proteins, and directs HLA class II molecules to endosomal compartments for ligand capture (9,10). After assembly, the HLA II-I chain complexes leave the ER and transit to mature endosomes either via the trans-Golgi network or by recycling from the cell surface (11). In acidic endosomes, I chain is progressively proteolyzed to the residual fragment, named class II-associated invariant chain peptide (CLIP), a variably extended peptide of roughly 20 residues, which remains bonded to the HLA-II binding groove (12). CLIP is subsequently released by the interaction with MHC-II-related glycoprotein, DM (HLA-DM), which further influences peptide selection in a process termed “peptide editing” by inducing the dissociation of low-affinity peptides. (9,13). The function of HLA-DM is modulated by another MHC-II-like heterodimeric protein, DO, and in vitro evidence most often support the view of DO as a negative modulator of DM (13). The inhibition of DM mediated by DO is strong at early endosomal pH (6.0–6.5) and attenuated at the lower pH (4.5–5.0) of late endosomal/lysosomal MHC II containing compartments (13). It has been suggested that DO modifies DM-mediated peptide exchange by limiting the location of fully functional DM, decreasing the presentation of antigens taken up by endocytosis and enhancing B cell receptor (BCR) mediated antigen presentation (13). After CLIP removal, the HLA class II molecules bind antigenic peptides. The HLA-II-peptide complexes are then transported to the cell surface and recognized by CD4+ T-cells to further the immune response mainly by providing helper signals for other immune cells important for the maintenance of self-tolerance and for the induction and regulation of adaptive immune responses against pathogens.

The peptide-binding site of the HLA class II molecules consists of a groove containing a flat part (floor) and of two walls formed of two helices (14). The floor is provided by an eight-stranded anti-parallel β sheet oriented away from the cell membrane and formed of $\alpha 1$ and $\beta 1$ domains. Each domain supplies four strands of amino acids (14). The peptide-binding grooves of HLA molecules are extremely polymorphic and contain hypervariable amino acidic residues located mainly in small cavities in the groove’s floor, called pockets (14). In addition, there are also clusters of highly conserved residues

between molecules of the same class, distributed along the length of the groove in class II molecules (14).

The nature of HLA-peptide binding is complex. MHC class II binding grooves are able to accommodate peptides of 10-20 amino acids length (Figure 1.4) (10, 15).

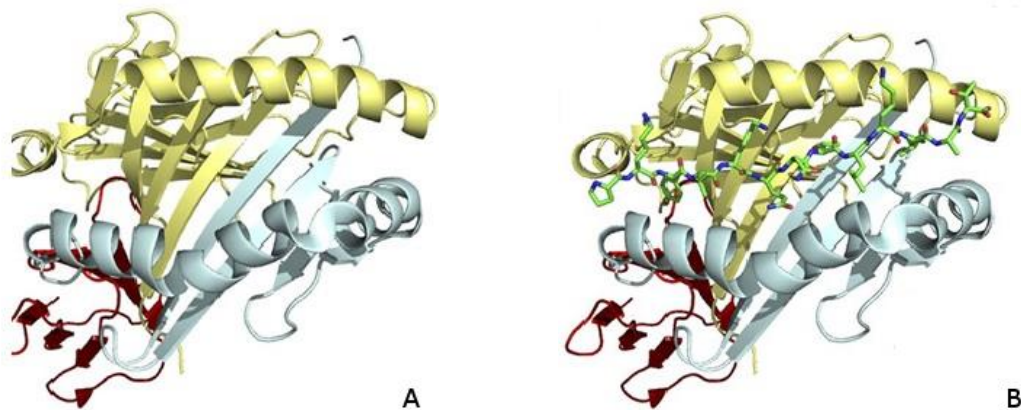


Figure 1.4 - Peptide binding groove of HLA-DRB4 molecule (A) and structure of HLA-DR4 with a peptide derived from influenza hemagglutinin, bound in its peptide-binding groove (B) (adapted from reference 15).

The peptides are bound in the groove in a largely extended conformation, which is independent of peptide's sequence and is not related to the conformation of the epitope sequence within the native antigenic protein (16). The conserved cleft side residues interact with the charged and polar residues of the main chain of a bound peptide, capable of adopting the required backbone conformation, through hydrogen bonds (17). These connections are named sequence-independent peptide bindings and contribute to the overall peptide-binding affinity (18).

In addition to the bonds with the peptide main chain HLA molecules make many contacts with side chains of bound peptide. The extended conformation of the bound peptide makes that the side chains of peptide residues at certain positions are directed into the pockets of the peptide binding groove (19). The peptide residues, which side chains interact with the side chains of HLA pockets, are termed anchor residues. These interactions increase the overall binding affinity, stabilize the bond between peptide and

HLA, and align the peptide in the groove so that its sequence can be read by the T-cell receptor (TCR) (18). The HLA class II molecules have several pockets, five of which accommodate side chains of the bound peptide (P1, P4, P6/7 and P9) (20). These pockets are named for the residue in the peptide sequence that interacts with them (the P9 pocket interacts with the 9th peptide side chain) (21). Deeper P1, P4, P6 and P9 pockets completely bury the peptide side chains whilst P7 is directed sideways within the groove. Since the pockets are formed of side chains of highly polymorphic amino-acidic residues of HLA molecule, their characteristics (such as volume, hydrophobicity and electrostatic charge) can alter the size, shape and charge of the pockets and their binding specificity (8). Therefore, the anchor-pocket interactions impose sequence requirements on the peptides that can bind (sequence dependent peptide bindings) (14). The sum of binding requirements at each peptide position defines the peptide sequence specificity, or “motif” of each HLA protein. It is determined by the unique combination of polymorphic residues in the peptide binding-groove, and therefore differs between HLA molecules derived from different alleles (22).

HLA-peptide binding is characterized by high-affinity but low- specificity. A single HLA molecule able to bind hundreds or thousands of different peptides that may have different sequence, but share amino acid residues (motifs) that fit into the anchoring pockets (anchoring residues). However, the high affinity interaction ensures that the same peptide remain bound for the life of HLA molecule.

HLA class II molecules present a peptide for the binding to TCRs of T CD4+ cells.

TCR is composed of two chains, α and β and contacts HLA-peptide complexes through the 6 complementarity-determining regions (CDR), three each from each chain (23). TCR align diagonally over the peptide; the peptide-binding groove and the less variable CDR1 and CDR2 loops interact primarily with the relatively conserved amino acid residues of the HLA α helixes, whereas the CDR3 loops engage the peptide presented in the HLA groove (Figure 1.5) (10, 23).

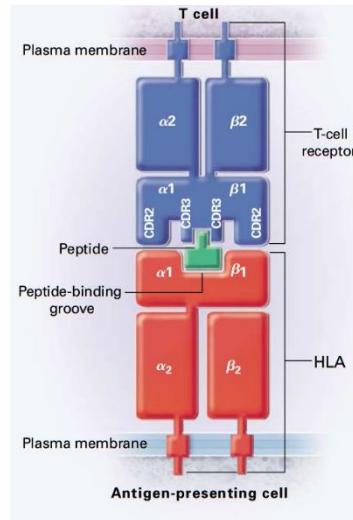


Figure 1.5 - Interaction between TCR and HLA-peptide complex. Peptide bond to the groove of HLA class II molecule interacts with CDRs of TCR (adapted from reference 10).

In that way, the contacts of CDR1 and CDR2 loops with the HLA molecule establish the overall orientation and configuration of the interaction, whereas the highly variable contacts of the CDR3 loops with the peptide determine the specificity of the interaction (23). There is the additional interaction between T cells and HLA molecules through binding of the CD4 co-receptor to the hydrophobic cleft between $\alpha 2$ and $\beta 2$ domains of HLA II molecule. This interaction is necessary for T cell activation and the main function of CD4 is to recruit and deliver lymphocyte-specific protein tyrosine kinase (Lck) to the TCR complex for initiation of T cell activation, proliferation, and differentiation signaling (24,25).

The precise molecular interactions by which a specific TCR will dock onto a peptide-HLA complex are currently impossible to predict, due partially to changes that occur at the protein-HLA interface after binding, (26,27). Flexibility at the interface, typically within the CDR loops, but also at the peptide and, less frequently, the HLA, plays a prominent role in binding. This adjustment of the components at the interface, named 'conformational melding', is an important feature of TCR recognition (28).

1.1.3. HLA class II diversity

HLA genomic region is one of the most gene-dense and polymorphic regions within the human genome (2). The extensive polymorphism has evolved in order to enable efficient binding of different peptides from the vast range of potentially pathogenic organisms (21). This evolutionary pressure varies with time and geographical area; therefore HLA has adapted differently in various populations and ethnical groups, creating different HLA phenotypes (29).

The variety of the HLA molecules is related to the extreme polymorphism of the HLA alleles and the codominant expression of HLA gene products (30). The diversity of HLA genes is reflected by the number of their alleles. According to the international ImMunoGeneTics HLA database (www.ebi.ac.uk/ipd/imgt/hla/stats.html) >14 000 class I and II HLA alleles have been recorded by March 2016 (Table 1.1) (31).

Table 1.1 HLA alleles (adapted from reference 31)

Numbers of HLA Alleles											
HLA Class I Alleles										10,730	
HLA Class II Alleles										3,743	
HLA Alleles										14,473	
Other non-HLA Alleles										177	
Number of Confidential Alleles										7	
HLA Class I											
Gene	A	B	C	E	F	G					
Alleles	3,399	4,242	2,950	21	22	53					
Proteins	2,396	3,131	2,089	8	4	18					
Nulls	157	134	109	1	0	2					
HLA Class I - Pseudogenes											
Gene	H	J	K	L	P	T	U	V	W	X	Y
Alleles	12	9	6	5	5	0	0	3	0	0	3
Proteins	0	0	0	0	0	0	0	0	0	0	0
Nulls	0	0	0	0	0	0	0	0	0	0	0

HLA Class II											
Gene	DRA	DRB	DQA1	DQB1	DPA1	DPB1	DMA	DMB	DOA	DOB	
Alleles	7	2,018	69	911	43	644	6	7	13	12	
Proteins	2	1,470	33	622	21	529	0	4	7	3	
Nulls	0	51	1	24	0	16	0	0	0	1	
HLA Class II - DRB Alleles											
Gene	DRB1	DRB2	DRB3	DRB4	DRB5	DRB6	DRB7	DRB8	DRB9		
Alleles	1,883	1	77	24	26	3	2	1	1		
Proteins	1,369	0	63	16	22	0	0	0	0		
Nulls	44	0	1	4	2	0	0	0	0		
Other non-HLA Genes											
Gene	HFE	MICA	MICB	TAP1	TAP2						
Alleles	6	105	42	12	12						
Proteins	4	82	28	6	5						
Nulls	0	2	2	1	0						
Component Information											
Number of Component Entries (EMBL/GenBank/DDBJ)										30,429	
Number of Source Cells/Individuals										26,244	

The particular combination of HLA alleles found on a single chromosome is known as an HLA haplotype (8). There is marked linkage disequilibrium in the HLA region and HLA haplotypes are often inherited “en block” (32).

A number of such haplotypes are present at a relatively high allele frequency, which may suggest a potential selective advantage (33). However, a possession of some haplotypes appears to be predominantly harmful and associated with a number of autoimmune diseases (2, 34).

The polymorphisms of the HLA genes are almost exclusively located in the gene regions which encode for the peptide-binding groove, and more precisely for the binding pockets,

and are situated in the exon 2 of the respective A and B genes (**5, 30**). The polymorphic differences in the variable regions of these exons are rarely allele-specific. Instead, the sequence motifs are often shared between a number of alleles (**30**).

The DR genes consist of multiple, highly polymorphic genes encoding for the β chain of HLA class II molecule (DRB genes) and scarcely polymorphic DRA gene codifying for invariant α chain (**31**). The HLA-DRB loci are variable in number, being the DRB1 the most variable, and are haplotype-dependent (**5, 35**). There are five different human DR haplotypes that comprise different combinations of DRB genes (**35**). Both DQA and DQB genes are polymorphic, as well as DPA and DPB genes. However, there is a striking difference in the allelic diversity of DQB and DPB alleles compared to DQA and DPA alleles respectively.

Accordingly, the amino acidic polymorphic residues are located mainly in the β chains of HLA class II molecules, clustered within the regions known as hypervariable regions (HVRs). HVRs are situated within the pockets of the HLA molecule peptide-binding groove. Genetic polymorphisms directly influence the conformation of the groove, and the set of peptides that can be presented by the HLA molecules.

The positions of HVRs of HLA class II molecules are the following:

- HLA-DRB: Only four genomic DRB loci are expressed, out of nine (DRB1, DRB3, DRB4, and DRB5 gene) and almost all of the DRB polymorphism is found at DRB1 (**31**). On DRB1, there are three HVRs, located at positions 9-13, 25-28 and 67-74 (**17, 36, 37**).
- HLA-DQ: There are three HVRs in the DQ β chain and they occur at positions 26-37, 52-57 and 70-74. The HVR in the DQ α chain has been identified at positions 40-56 (**38, 39**).
- HLA-DP: The five HVRs in the β chain include amino acids at positions 8-11, 35-36, 55-57, 65-69 and 84-88 (**39, 40**). The polymorphism of the α chain is extremely limited. The amino acid at position 31 participates in the P1 binding pocket, but is located on the floor of the peptide-binding groove and is not exposed to the TCR (**41**).

Additional diversity arises from the ability of α and β chains to mix and match with chains expressed from both parental alleles. However, there is the evidence that some of these pairings do not produce stable surface heterodimers, and thus certain combinations of DA and DB alleles are almost never seen on the same haplotype (**41, 41**).

1.1.4. Shared epitope hypothesis

The shared epitope hypothesis is based on the assumption that HLA class II molecules are directly involved in the etiopathogenesis of autoimmune diseases and that structural differences between class II molecules can influence their interaction with T cell receptor and/or antigens.

This hypothesis was developed in the field of rheumatoid arthritis (RA). The investigations on the relation of MHC alleles and susceptibility to RA started following the observation by Astorga and coll. that mixed lymphocyte culture (MLC) reactions, a test used for the detection of histo-incompatibility in vitro, between the lymphocytes of patients with RA were frequently mutually non-stimulatory (**43**). Stastny first defined the genetic basis for this phenomenon by demonstrating an elevated frequency of the MLC type DRw4 (old denomination for HLA-DRB1*04 according to the new nomenclature) in RA patients (**44**). Subsequently, further findings demonstrated that some of the genes encoding DR4 serologic specificity were primarily associated with disease susceptibility (**45-47**). Nevertheless, other alleles, notably DR1 (now HLA-DRB1*01:01), were found to be associated with RA in patients who lack DR4 haplotypes (**48,49**). In order to explain this discrepancy, and to clarify why some of the DR04 subtypes confer risk for RA and others not, Gregersen et coll. hypothesized the existence of the common, shared, amino-acid sequence, codified by different HLA-DRB1 alleles associated with RA in different ethnical populations and known at that time (**50**). The authors noted a striking sequence homology within the third hypervariable region of alleles: DR1, Dw4, Dw14, Dw15 and Dw16 (now: DRB1*01:01, DRB1*04:01, DRB1*04:04 and DRB1*04:05, according to the new nomenclature) that codified for amino acids at

positions from 67 to 74, and hypothesized that this “shared epitope” or group of related epitopes may fold into conformationally equivalent structures and act as potentially functional units (**50, 51**). In this manner, what is shared among the HLA molecules is a conformation capable of specific binding. These functional units would bind the same peptide(s) and/or would interact with particular subpopulation of T cells, thus triggering an auto-immune response resulting in RA. In addition, this finding explained why the disease association studies based solely on conventional serologic analysis failed to clearly explain the association of RA and HLA until then.

Later interpretation of this hypothesis led to the concept that susceptibility to RA is determined by the amino acid sequences QKRRAA and QRRAA, as well as by the RRRAA, at positions 70-74 (**52**). All of them contain a common motif, (shared epitope) represented by RAA (arginine, alanine and alanine) at positions 72–74. The alleles that encodes for proteins that include the shared epitope (SE) sequences are also called “shared alleles”.

However, not all SE alleles display the same magnitude of association with RA. DRB1*04:01 and *04:04 confer a stronger predisposition to RA than DRB1*01:01 and *10:01 (**53, 54**). Moreover, particular combinations, such as DRB1*04:01/*04:04, appear to confer a particularly high risk of RA (**53, 55**).

Moreover, some of the HLA-DRB1 alleles display protective effect in respect to RA.

Zanelli et al. discovered that the DRB1 alleles that encode for HLA proteins carrying DERAA motif at position 70-74 (alleles DRB1*01:03, *04:02, *11:02, *11:03, *13:01, *13:02) may be protective towards RA (**56**). The protective role of DERAA positive alleles, even in the presence of alleles carrying the shared epitope, was confirmed by further studies (**57, 58**). In order to further explain the influence of HLA-DRB1 alleles on RA susceptibility, not fully explained by the SE hypothesis, and to include the protective alleles, as well as the protective effect of isoleucine (I) at position 67, de Vries et al. extended the shared epitope to include amino acids 67–74 (**58**). Matthey et al. added a protective role of aspartic acid (D) at position 70 (**59**). Subsequently, du Montcel et al. suggested a new SE model and classification in RA (**60**). They proposed that RA susceptibility was conferred by RAA at positions 72–74, as previously

hypothesized for the SE, but that this was modulated by amino acids at positions 70 and 71 (60). The concurrent presence of lysine (K) at position 71 to RAA conferred the highest risk for RA, arginine (R) was associated with intermediate risk, and alanine (A) or glutamic acid (E) with the lowest risk. Glutamine (Q) or arginine (R) at position 70 conferred greater risk than aspartic acid (D). The shared alleles were divided into five groups according to amino acids at positions 70 and 71. Since the three of the groups carried approximately the same risk for RA, the groups become three: S2 (XKRAA), S3P (Q/RRRAA) and L (low risk alleles). This allelic classification has been validated in different ethnical populations (61, 62) and the large UK study that included more than 1300 RA patients, confirmed the association of S2 and S3P with RA and validated the hierarchy of risk proposed by this classification (63).

Furthermore, a recent study performed on 5000 ACPA (anticitrullinated peptide antibodies) positive RA cases and 14900 controls of European descent from six independent genome-wide data sets, suggested that a model including specific amino acids at positions 11, 13, 71 and 74 of the HLA-DRB1 chain, as well as at position 9 of HLA-DQB1 and 9 of HLA-B, almost completely explain the HLA association with RA risk (54).

The authors replicated HLA- DRB1 effects on sequencing-based classical HLA-DRB1 genotypes at 4-digit resolution in an independent South Korean cohort of 616 ACPA positive cases and 675 controls (54). Amino acid position 11 demonstrated the strongest association; the aliphatic residues Val-11 and Leu-11 conferred the highest risk, whereas the polar Ser-11 residue was highly protective against the disease. All effects were conditional on Asp-9 in HLA-B and Phe-9 in HLA-DPB1 (54).

Besides being associated to RA susceptibility, SE is correlated with disease severity and bone damage (64,65). Associations of extra-articular manifestations with the SE or DRB1*04 alleles have also been reported (66).

Moreover, a gene dosage effect has been noted for the association of SE alleles and RA. Therefore, having two SE alleles confers a higher risk for RA than having one allele (67). In order to explain this effect, it has been hypothesized that SE molecules may act as allele-specific signal transduction ligands that interact with non-HLA cell surface

receptors, including cell surface calreticulin, for production of nitric oxide and reactive oxygen species for cell signalling (68).

SE alleles are mainly associated with ACPA positive RA (69-71). ACPA, present in around 70% of RA patients, (72), are highly specific for RA and are associated with more severe joint damage and radiographic outcome (72-74). SE seems to influence ACPA levels as well as their fine specificity (75, 76). On the other hand, only a weak association of SE alleles with RF and ACPA negative RA has been reported (77, 78). ACPA are believed to have a role in the pathogenesis of RA and their presence is considered a predictive marker of bone destruction in RA (79). It is likely that the SE itself is not directly responsible for poorer prognosis, but that might rather influence the outcome indirectly, via ACPA production (80).

A number of models have been proposed to explain the association and effect of SE and RA susceptibility. It has been thought that shared alleles conferred disease susceptibility through the selective presentation of self-peptides to T cells (81) or that shared alleles may shape the peripheral T-cell repertoire by presenting self peptides to CD4 T cells in the thymus (82).

At present, it seems that HLA-DRB1 molecules containing SE display a peptide binding pockets with a high affinity for citrullinated antigens, which are found at higher levels in subjects prone to develop RA.

Four residues that are hypothesized to be a part of the broad SE (67,70,71 and 74) contribute to the P4 pocket of the peptide binding groove of the β -chain (37). The lysin (K) or arginin (R) at position 71 are positively charged and make that this pocket has high affinity for negatively charged or uncharged polar residues (19).

The amino acid residues at positions 11 and 13 of HLA-DR β , which also predispose strongly to RA, form the binding pocket for peptide position 6 of DR molecules (37). Furthermore, the positions 71 and 74 are separated by a single turn along the α -helix, and their side chains are spatially close to those of positions 11 and 13 (54).

After the discovery of antibodies against citrullinated peptides, the studies on HLA molecules carrying SE showed that the conversion of arginine to citrulline, a process which replaces the charged imino side-chain group with an uncharged carbonyl,

dramatically increased the affinity of the vimentin peptide for SE + DRB1*0401 through increased affinity for P4 (**83, 84**).

A structural study confirmed that HLA-DRB1*04:01 and HLA-DRB1:*04:04 proteins presented citrullinated vimentin and aggrecan peptides through interactions between lysine or arginine at position 71 of DRB1- β chain and citrulline at the P4-binding pocket of DRB1 molecules (**85**).

The critical roles for Lys71 β and Arg71 β in the presentation of citrullinated peptides are consistent with the findings that DRB1 alleles that carry glutamic acid at position 71 of the β chain are not associated with RA risk (**86**). Regarding the amino acids at positions 11 and 13 of β chain of HLA DRB1 molecules which also predispose strongly to RA, self-peptide(s) that might explain the predispositions of 11 β and 13 β to RA have not currently been identified.

SE alleles exert pathogenic effects through the presentation of citrullinated peptides, which are recognized as non-self by T cells, but the exact mechanisms underlying the effect of the SE in RA has not been elucidated. It has been suggested that the central and/or peripheral tolerance to post-translational modified citrullinated peptides may be incomplete (**87**). Therefore, citrullinated proteins could be taken up by antigen-presenting cells in inflamed tissues, where citrullination occurs, and presented by DRB1 SE molecules to CD4 T lymphocytes. T cells could then provide help to B lymphocytes, resulting in the production of ACPA antibodies (**84, 88**).

Another recent hypothesis is that SE may contribute to the genetic predisposition to RA causing an immune dysregulation (**89**) or a premature immunosenescence (**90**).

1.2. Systemic sclerosis and chronic sclerotic graft versus host disease

1.2.1. Systemic sclerosis

Systemic sclerosis (SSc) or scleroderma is a severe autoimmune connective tissue disease that affects skin and internal organs. The scleroderma (tight and hard skin) is also the hallmark of disease. SSc can be divided into diffuse cutaneous SSc (dcSSc) and limited cutaneous form of SSc (lcSSc) according to the extent and distribution of skin involvement (91). In patients with limited cutaneous subset skin tightening is distal to the elbows and knees. In dSSc, the skin of the proximal extremities and trunk is also involved.

Principal clinical manifestations of SSc are summarized in Table 1.2.

Table 1.2 - Main clinical manifestations of SSc

Skin and vascular	- Raynaud's phenomenon - Puffy fingers - Skin thickening - Sclerodactily - Digital ulcers and skin ulcers in other sites - Gangrene and digital loss - Teleangiectasia - Calcinosis
Gastro-intestinal	- Microstomia - Xerostomia - Gastro-oesophageal reflux and oesophageal dysmotility - Delayed gastric emptying - Gastric antral vascular ectasia (GAVE) - Malabsorption syndrome

	- Small intestinal bacterial overgrowth - Impaired bowel transit and intestinal pseudo-obstruction - Anorectal dysfunction and faecal incontinency
Lung and heart	- Interstitial lung disease and lung fibrosis - Pulmonary arterial hypertension - Myocarditis and cardiomyopathy - Pericardial effusion - Ventricular arrhythmia and sudden death
Kidney	- Scleroderma renal crisis - Impaired glomerular filtration rate
Musculoskeletal and peripheral nerve system	- Arthralgia - Arthritis and Tenosynovitis - Tendon friction rubs - Carpal tunnel syndrome - Myopathy - Myositis - Acroosteolysis - Osteopenia and osteoporosis

The prevalence of SSc has been estimated to be around 15-28/1 000 000, with the incidence of 1,5 - 2,5 million/year, although data may vary in different geographic areas (92). The mean onset of disease is from 40 to 70 years, and women are more frequently affected than men (93). The estimated cumulative survival from the diagnosis is approximately 75% at 5 years, 60% at 10 years, and around 40% at 20 years. (94) and main cause of death is lung involvement and in particular pulmonary arterial hypertension and pulmonary fibrosis (94-96).

The pathogenesis of SSc involves interplay between: a) innate/adaptive immune system abnormalities leading to production of autoantibodies and cell-mediated autoimmunity, b) microvascular endothelial cell/small vessel fibroproliferative vasculopathy, and c) fibroblast dysfunction generating excessive accumulation of collagen and other matrix components in skin and internal organs (**97, 98**). This pathogenetic scenario leads to microvascular damage, tissue inflammation and fibrosis inducing eventually skin fibrosis and internal organ impairment. The disease is heterogeneous in its clinical presentation that likely reflects different genetic or triggering factor influences on the immune system, vasculature, and connective tissue cells.

1.2.2. Pathogenesis of fibrosis in SSc

Excessive deposition of the components of extracellular matrix, including collagens, proteoglycans, fibronectin and hyaluronic acid and subsequent fibrosis of skin and internal organs is one of the main causes of tissue damage and disease progression in SSc, and it has been correlated with disease severity (**99**). Once initiated, fibrosis can escalate through multiple amplification loops that are generated as a consequence of tissue damage, increased matrix stiffness, hypoxia and oxidative stress, creating a 'vicious' cycles of profibrotic events (**100, 101**). Progressive fibrosis ultimately leads to organ failure and substantial mortality, and to date has no effective therapy.

In SSc the autoimmunity and vasculopathy characteristically precede fibrosis. The current concept is that fibrosis in SSc begins with an initial vascular and/or immune tissue insult (**102**).

Microvascular dysfunction consisting in impaired vascular tone and vascular permeability, and endothelial cell injury and activation are one of the earliest and possibly primary pathogenic events in course of SSc (**103, 104**). Endothelial injury and activation can be caused by: viral agents, ischemia/reperfusion injury, cytotoxic T cells, inflammatory cytokines and antibodies, but the exact etiological mechanism is still unknown (**103-105**). Activated endothelium up-regulates the expression of vascular and intracellular adhesion molecules, allowing the attachment of platelets and leucocytes,

and releases vasoactive mediators, such as endothelin-1, a potent vasoconstrictor that also promotes vascular smooth muscle cell proliferation and fibroblast activation (**97, 100, 105, 106**). Damaged endothelium activates macrophages, monocytes, T lymphocytes, platelets and fibroblasts to aberrantly secrete growth factors, cytokines and chemokines, inflammatory mediators, vasoactive and profibrotic agents, and drives chronic inflammation, tissue damage and wound-healing responses.

Transforming growth factor- β (TGF- β) has been considered the master regulator of both physiological and pathological fibrogenesis (**103**). It is secreted by macrophages (main producers), monocytes, platelets, T lymphocytes and fibroblasts (**107**) and the responses to TGF- β are context and target cell lineage dependent. In mesenchymal cells, it acts as a potent fibrogenic stimulus by enhancing collagen synthesis, proliferation, migration, adhesion, and trans-differentiation into myofibroblasts (**103**).

Other molecules that mediate fibrosis are represented by connective-tissue growth factor (CTGF) (**108**), platelet-derived growth factor (PDGF) (**109**), Interleukin 13 (IL-13), IL-4 (**110, 111**), IL-6 (**112**), platelet factor 4 (known also as CXCL4: chemokine (C-X-C motif) ligand 4) (**113**), macrophage inflammatory protein MIP-1 (known also as CCL2: chemokine (C-C motif) ligand 2) (**114**), monocyte chemoattractant protein 1 and 3 (**115**), lysophosphatidic acid (LPA) and sphingosine-1-phosphate (S1P) (**116**). In addition, Wnt- β -catenin (**117**) and Notch pathways (**118**), erythropoietin-producing human hepatocellular carcinoma (ephrin) (**119**) and integrins (**120**) are also critical for fibrotic process in SSc. These mediators drive inflammation and fibroblasts activation and recruitment of progenitor and/or stem cells from the bone marrow and circulation (**100**) and are implicated in the trans-differentiation of stromal cell types, including pericytes, epithelial cells, endothelial cells, and adipocytes, into apoptosis-resistant profibrotic myofibroblasts (**97**).

Innate and adaptive arms of the immune system are also involved in the initiation of SSc and in the fibrogenesis in course of the disease. The immune system can be activated as a consequence of vascular damage or can be the cause of vascular injury through the production of anti-endothelial antibodies or by cytotoxic T cells and antibody-dependent cellular cytotoxicity (**104**). Regardless of the initiating stimulus, chronic tissue

injury results in the activation of the innate immune responses controlled by monocyte derived cells (100). The cells of the innate immune system, both resident in the tissues or derived from the inflammatory infiltration, recognise the endogenous ligands, such as damage associated molecular patterns, endogenous proteins, RNA, DNA and other cellular products through Toll-Like Receptors (TLR) and trigger the innate immune signalling and gene activation (121). Innate immune signalling via TLRs is increasingly recognized as a key player driving the persistent fibrotic response in SSc (101). Ligand-induced TLR4 activation seems to exert a potent stimulatory effect on fibrotic gene expression and myofibroblast differentiation. Furthermore, TLR4 appears to sensitize fibroblasts to the profibrotic stimulatory effect of TGF- β (101).

In the non-lesional skin of SSc patients there is a perivascular infiltrate consisting of macrophages/monocytes and CD4⁺ T cells supporting the hypothesis that the alterations of immune system precede fibrosis (122). Activated macrophages and monocytes are able to produce TGF- β 1 and pro-fibrotic agent tissue inhibitor of matrix metalloprotease-1 (122, 123). In SSc there seems to exist M2 macrophage polarization which participates to pathogenic mechanisms of skin and lung fibrosis (124). Activated macrophages can be divided in M2 or alternatively activated macrophages, commonly observed during the peak of the profibrotic immune response and involved in tissue remodelling and wound healing, and in classically activated macrophages, M1, which secrete high levels of pro-inflammatory cytokines and exert strong bactericidal and tumoricidal effects (125).

Type I interferons (IFNs) are important key regulators of the innate immune system. Deregulation in type I IFN and IFN-inducible genes are involved in pathogenesis of SSc and its fibrotic manifestations (122, 126).

Interleukin-6 (IL-6) is one of the most prominent cytokines activated by IFN pathway (122). It is a classic inflammatory cytokine produced by various cells and involved in induction of acute phase proteins, B cell proliferation and differentiation of T lymphocytes (127). IL-6 binds to its receptor (IL-6R) and activates downstream signalling mediated by signal transducer and activator of transcription (STAT)1 or STAT3 through

tyrosine phosphorylation (**128**). It has been shown that IL-6 plays a role in the initiation and promotion of fibrosis in SSc. (reviewed in **129** and **130**).

T cells play an important role in fibrosis in SSc through activation of monocytes and direct release of pro-fibrogenic cytokines and the main role is played by CD4⁺ helper T lymphocytes (**105**). There are two main types of T helper (Th) cells: Th1 cells and Th2 cells (**131**). Th1 cells produce cytokines, such as interferon γ (IFN γ), IL-1, IL-2, IL-12 and tumour necrosis factor (TNF), which suppress collagen production by fibroblasts in vitro, whilst Th2 produce cytokines, such as IL-4, IL-13 and IL-6, that stimulate the synthesis of collagen by human fibroblasts (**132**).

Increased levels of IL-4 and IL-13 have been observed in SSc patients (**111**). IL-4 and IL-13 have been shown to stimulate the transcription of the human $\alpha 2$ type I collagen gene (*COL1A2*) in dermal fibroblasts (**133**). IL-4 promotes fibroblast proliferation and collagen synthesis and stimulates the production of TGF- β (**134**). IL-13 may induce fibrosis through TGF- β -independent mechanisms, such as stimulating fibroblast proliferation and collagen production through the activation of innate immunity (**135**). Th17 lymphocytes are another type of Th cells characterized by the production of cytokines, such as IL-17 and IL-22 (**136**). Th17 cells are increased in blood, skin and lung tissues of SSc patients (**137, 138**). Studies in mice have shown that IL-17A may promote the production of TGF- β and collagen by skin fibroblasts (**139**). However, in humans TH17 cells inhibit collagen production in healthy fibroblast, whereas the SSc fibroblast are resistant to collagen inhibition mediated by Th17 cells (**140**).

There is evidence that early stages of disease are characterized by Th2/Th17 immune polarization and skin infiltration of Th2 and Th17 cells, macrophages and mast cells (**141-143**). It seems that there is a shift from Th2 to Th1 along with the disease duration and that the sustained Th2 response relates to the exacerbations of SSc (reviewed in **144**). Studies in mice model of SSc have shown that different cell adhesion molecules expressed on endothelial cells may regulate diverse T cell infiltration (**145**). E-selectin and P-selectin, regulate Th1 accumulation in the skin whilst ICAM-1 and L-selectin and ICAM-1 regulate Th2 and TH17 infiltration (**145**). Thus it has been hypothesized that the

infiltration of T cells is in SSc regulated by the altered expression of cell adhesion molecules **(144)**.

T regulatory cells are also involved in the pathogenesis of fibrosis in SSc. It has been reported that the number of Tregs is decreased in the skin of SSc patients **(146, 147)**, but that skin Tregs are able to produce pro-fibrotic cytokines IL-4 and IL13 **(148)**.

B lymphocytes may also contribute to fibrosis in SSc **(149)**. Activated B cells produce profibrotic cytokines IL-6, and TGF β **(149)**, and autoantibodies, some of which, such as anti-PDGF antibodies **(150)**, anti-matrix-metalloproteinase-1 **(151)** and anti-matrix metalloproteinase-3 antibodies **(152)** may induce collagen synthesis. Anti-PDGF antibodies recognize native PDGFR, induce tyrosine phosphorylation and accumulation of reactive oxygen species (ROS), and stimulate type 1 collagen gene expression **(150, 153)**. In addition, B cells can function as antigen-presenting cells to T lymphocytes and through cell contact may induce dendritic cell maturation that promotes profibrotic Th2 response **(154)**.

Fibroblasts and myofibroblasts are the principal effector cells of fibrosis **(155)**. Myofibroblasts are specialized activated fibroblasts which present features of smooth muscle cells, such as the expression of contractile protein alpha smooth muscle actin (α -SMA) **(156)**. They synthesize increased amounts of collagen (types I and III), other ECM components, and collagen-modifying enzymes **(97, 105, 157)** and are major responsables for the development and progression of fibrotic. In addition, myofibroblasts are a major source of TGF- β **(100)**. In SSc, myofibroblasts may derive from the expansion and activation of resident fibroblasts **(158)**, but also from other types of cells, such as bone marrow-derived fibrocytes **(155)**, epithelial cells through a process known as epithelial to mesenchymal transition **(159)**, pericytes **(160)**, adipocytes, macrophages and also endothelial cells through endothelial to mesenchymal transition **(161)**.

Fibroblasts isolated from affected tissues in SSc display an abnormal activated phenotype characterized by enhanced extracellular matrix synthesis, constitutive secretion of cytokines and chemokines, increased expression of cell surface receptors for fibrogenic signaling mediators and show evidence of myofibroblast trans-differentiation **(103, 162)**. Chronic inflammation and hypoxia with oxidative stress, the

presence of systemic and local profibrotic factors facilitate the trans-differentiation of resident fibroblasts and recruited cells in SSc (**100**). High ET-1, TGF- β , CTGF and FGF-2 and other growth factors and/or cytokines are mediators of the transformation into myofibroblasts (**163, 164**).

Recent studies indicate that the increased mechanical stress derived from the matrix remodelling and increased matrix stiffness may also contribute to the fibroblast activation (**165, 166**). Moreover, it has been hypothesized that SSc fibroblasts may have an increased susceptibility to the effects of the mechanical stimuli compared to healthy fibroblasts. (**167**).

As the final result of the fibrotic process, there is a progressive replacement of tissue architecture by components of the extracellular matrix, resulting in damage and functional impairment of affected organs. The main pathogenic pathway of fibrosis in SSc are reassumed in Figure 1.6.

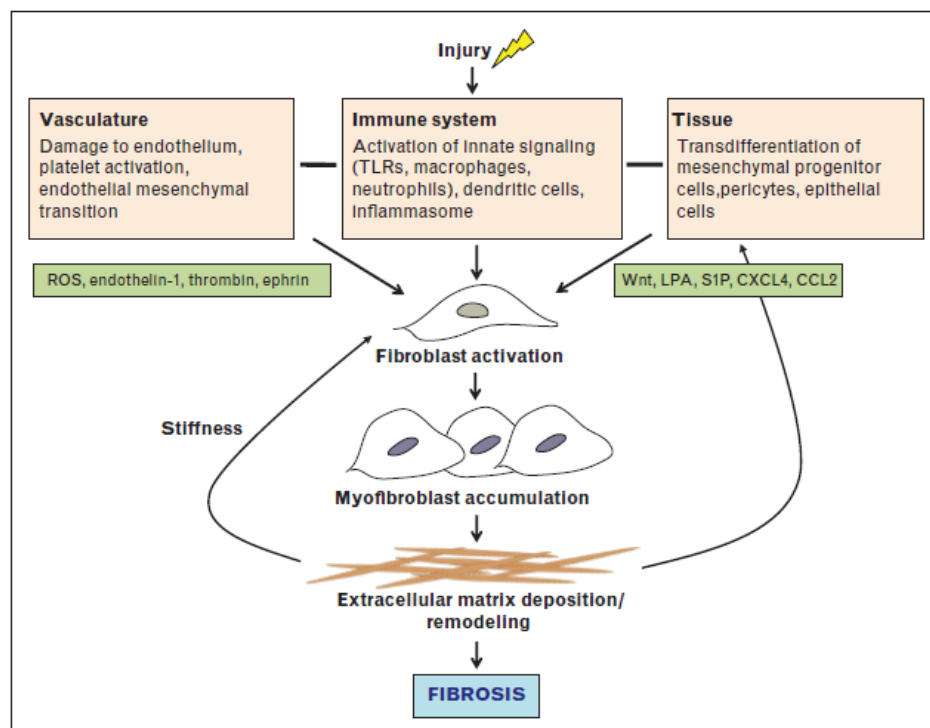


Figure 1.6 - Pathogenesis of fibrosis in systemic sclerosis. Endothelial damage and immune system activation leads to the release of inflammatory cytokines, growth factors, and ROS. Fibroblast activation and transdifferentiation of stromal cell types cause myofibroblast accumulation and deposition of extracellular matrix (adapted from reference 102).

1.2.3. Sclerotic graft versus host-disease

Chronic graft versus host disease (cGVHD) is one of the most severe complications of the allogeneic hematopoietic stem cell transplantation (HSCT) and is currently the main cause of long-term morbidity and mortality following HSCT (**168, 169**). Its clinical characteristics resemble autoimmune and other immunologic disorders such as SSc, Sjögren syndrome, primary biliary cirrhosis, bronchiolitis obliterans, immune cytopenias and chronic immunodeficiency (**170**). Sclerotic (sclerodermatous) GVHD (sclGVHD) is a distinctive clinical form of cGVHD with sclerotic features in the skin characterized by thickened, tight, and fragile skin (**170**). SclGVHD includes several cutaneous clinical pictures characterized by inflammation and progressive fibrosis of the dermis and subcutaneous tissues, resembling morphea, systemic sclerosis, or eosinophilic fasciitis (**171**). Skin manifestations comprise deep sclerotic features (e.g. smooth, waxy, indurated skin—"thickened or tight skin," caused by deep and diffuse sclerosis over a wide area) and morphea-like superficial sclerotic features (e.g. localized patchy areas of moveable smooth or shiny skin with a leathery-like consistency, often with dyspigmentation). (**170**). Eosinophilia is observed in about 50% sclGVHD patients (**172**). It has been estimated that sclGVHD occurs in around 13-20% of allogeneic HSCT recipients at 5 years after the transplant (**171, 173-175**). SclGVHD usually occurs late in the course of cGVHD, with a median onset of 15 - 16 months (**173, 174**).

Risk factors for sclGVHD are represented by young patient age (**175**), non-T cell depletion (**174**), multiple myeloma (**175**), peripheral blood stem cell as the stem cell source (**174, 175**) and total body irradiation in the conditioning regimen (**171, 173**). Anti-thymoglobulin in conditioning regimen and cord blood unit as the stem cell source are associated with protective effect against (sclGVHD) (**175**). ABO blood group mismatch and HLA mismatching seem also to exert a protective effect (**173**).

Although usually sclGVHD is not an acute life-threatening manifestation, severe sclGVHD may cause significant morbidity and disability because it is often associated with poor wound healing, inadequate lymphatic drainage, skin ulcers and joint contractures (**174**,

176). The response to topical interventions is poor and sclGvHD is often refractory to systemic therapy (**177**).

1.2.4. Pathogenesis of cGVHD and sclGVHD

GVHD is a complex immunological disorder in which donor-derived alloreactive T cells recognize the host antigens not expressed in the donor and initiate a tissue damage. Donor T cells recognize both host allogeneic histocompatibility antigens and unique tumour antigens exerting an anti-leucemic effect, through an immune-mediated graft versus tumour (GVT) effect (**178**). Classically, three requirements for the development of GVHD have been postulated: the graft must contain immunologically competent T cells; the recipient must express tissue antigens which are not present in the transplant donor; and the recipient must be incapable of establishing of an effective response against the transplanted cells (**179**). There are two forms of GVHD: acute and chronic GVHD, involving distinct pathological processes. Acute GVHD (aGVHD) has strong inflammatory components and cGVHD display more autoimmune and fibrotic features (**178**). CGVHD can occur as an extension of aGVHD, with its onset during or directly after an episode of aGVHD. It can also occur after an aGVHD free interval or it can present de novo, without previous aGVHD (**180**).

AGVHD is mainly due to the transfer of mature T cells whereas; in contrast, cGVHD is thought to be caused by donor cells that develop de novo within the host from the transplanted haematopoietic stem cells (**181**).

In aGVHD, recipient's antigen-presenting cells (APCs), present alloantigens (host HLA proteins and minor histocompatibility antigens) to donor alloreactive cytotoxic T cells leading to their activation that leads to the tissue damage. This process is amplified by pro-inflammatory cytokines released from tissues damaged by conditioning, and the entrance of lipopolysaccharides and other pathogen-associated molecules through damaged gut mucosa, which in turn up-regulate the innate immune system (**182**).

The knowledge on pathogenesis of cGVHD and sclGVHD is poorly understood and derives mainly from the animal models, which however do not reproduce the full clinical spectrum seen in humans. Schroeder et al have comprehensively resumed the main pathogenic mechanisms of cGVHD and sclGVHD in animal models (183) (Figure 1.7), although more recent findings have elucidated some of the pathogenic aspects of cGVHD.

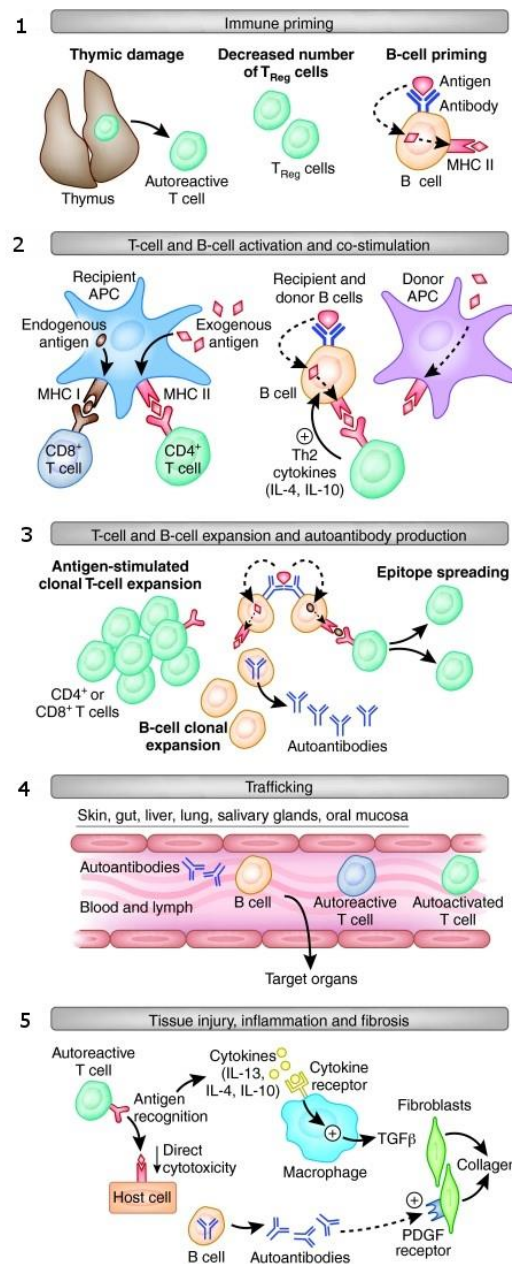


Figure 1.7 Steps leading to cGVHD and sclGVHD in mice (adapted from reference 183) (see text).

Differently to aGvHD, immune priming, at least in cGvHD murine models, is not dependent on tissue damage from conditioning. Compared with aGvHD models, the initial inflammatory environment is reduced or absent in cGvHD models, which involve reduced levels of radiation or no conditioning at all (**183**).

There is a thymic damage which can result in the impaired negative selection of autoreactive T cells (**184, 185**). This together with a decrease in the number or function of T regulatory cells (**186**) may contribute to the breakage of self-tolerance. Additionally, recipient's and donor's B lymphocytes recognize and process extracellular antigens for presentation in the context of MHC (**187**). In mouse models, T cell activation does not depend on recipient's APCs, but is a mixture of cognate antigen recognition and cross-priming of donor T cells (**183**). T CD4⁺ helper lymphocytes are necessary and sufficient for the development of cGVHD in murine models, whilst T CD8⁺ cytotoxic lymphocytes are not required, differently from sGVHD (**188**). The current concept is that the donor's T cells are essential for the development of GVHD (**178**). However, Lee et al. have shown that that cGVHD may be caused by autoreactive host T cells in mice that have been lethally irradiated and grafted with T-cell depleted allogeneic bone marrow cells (**189**). Recipients B cells are activated by minor antigens and cytokines, and mature to present processed antigens to donor CD4⁺ T cells, which in turn co-stimulate B cells to produce autoantibodies and additional stimulatory cytokines (**178, 183, 187**). Subsequently T and B lymphocytes clonally expand in response to antigenic stimulation and co-stimulation to form memory and effector cells (**183**). Expanded T and B cells can move to target organ sites and cause damage through cytokine-mediated fibrosis and chronic inflammation (**178, 180, 190**). In addition, bone-marrow-derived T cells which were inappropriately selected in a dysfunctional thymus, or by impaired peripheral tolerance, reach target organs where can be activated by self antigens and perpetuate the Th2 response (**180, 191**). Recent studies have also suggested the role of Th1 and Th17 cells in the pathogenesis of cGVHD and sclGVHD (**192, 193**). Eventually, tissue injury results from activated T and B cells that release inflammatory and fibrosing cytokines, such as IL-4, IL-13 and IL-10, and activated macrophages producing transforming growth factor- β (TGF β) stimulate collagen production from fibroblasts, collectively leading to a SSc-like

syndrome (**180, 190**). Autoantibodies may also be deposited at the site of target organs, causing activation of the platelet derived growth factor receptor, leading to fibrosis (**153**).

In detail, three main pathological mechanisms have been hypothesized in the pathogenesis of cGVHD: a) breakage of self tolerance and defects in thymic function b) B cell deregulation and autoantibody production and c) fibrosis (**180**).

The immune tolerance to self-antigens is disrupted in cGVHD, causing autoimmune manifestations of the disorder. It has been hypothesized that the thymic damage that is caused by conditioning regimens and/or acute GVHD leads to deregulation of central tolerance mechanisms during the reconstitution of the immune system post-transplantation (**194**). Sakoda et al. showed that impaired thymic negative selection of the recipients allowed the emergence of autoreactive T cells and caused cGVHD in a murine model characterized defined as sclerodermatous skin change and/or one of the following pathologic features: skin sclerotic change, ductopenia/portal fibrosis in the liver, or destruction/fibrosis of the salivary glands (**195**). Thymectomy prevented cGVHD in this model (**195**). On the other side, Zhang C et al showed that in the murine model made by the transplantation of DBA/2 spleen cells into sublethally irradiated BALB/c mice, the host thymus was not required for the induction of cGVHD and that quiescent autoreactive T and B cells in transplants from non-autoimmune donors might be activated and expanded to cause cGVHD (**188**). In cGVHD derived from aGVHD, the disease was caused by host reactive donor CD4⁺ cells generated via thymic dependent pathways in recipients with donor CD8⁺ T cell-mediated aGVHD. The authors have hypothesized that damage to the thymus mediated by alloreactive CD8⁺ T cells might result in the aberrant recovery of immune function and dysregulation of central tolerance, leading to the development of pathogenic, GVHD-inducing CD4⁺ T cells (**184**). Subsequently, it has been reported that mice with aGVHD have an important decrease in the mature medullary thymic epithelial cells (mTEC) expressing the autoimmune regulator (Aire), being targets of donor T cell alloimmunity (**185, 196**) and that thymic aGVHD interfered with the capacity of these cells to sustain tissue-restricted antigens (TRA) diversity (**185**). The diverse expression of TRA by these cells is required for the

effective thymic elimination of auto-reactive T lymphocytes by negative selection and since these cells are necessary for the thymic production of naturally occurring Treg lymphocytes (197). Moreover, the expression and diversity was selectively decreased in the tissues that are the target organs of human cGVHD (185). Dertschnig hypothesized that a decline in the pool size of Aire+mTEC which eliminates individual tissue-restricted peripheral self-antigens from the total thymic ectopic TRA repertoire, weakens the platform for central tolerance induction and favour the development of cGVHD, thus linking acute and cGVHD (198). However, the studies on sclGVHD in humans showed that aGVHD is not a risk factor for the occurrence of sclGVHD (171, 173-175).

Patients with cGVHD have lower frequency of CD4 Treg cells when compared with patients without cGVHD (199) and it has been suggested that CD4Treg deficiency could enhance alloreactivity and promote cGVHD (200). On the contrary, prompt recovery of CD4Tregs can prevent cGVHD while also supporting recovery of a broad T-cell repertoire (186). Alho et al. have recently examined reconstitution of CD4Tregs, CD4Tcons, and CD8 T cells, in prospectively monitored cohort of 107 adult patients who underwent allogeneic HSCT. Patients which developed cGVHD were characterized by a lower percentage of CD4Tregs due to the significantly inferior proliferation of central memory (CM) CD4Tregs (201). Early recovery of naive and CM fractions within each T-cell population 3 months after transplant was also strongly correlated with the subsequent development of cGVHD (201). This last finding was already reported in murine models which showed that T cells capable of promoting GVHD are primarily contained within the naive T-cell compartment (202). Moreover, a recent clinical trial demonstrated that depletion of naive T cells from the stem cell product markedly reduced the incidence of cGVHD, whilst had little effect on aGVHD (203). The authors have hypothesized that the increased expansion of relatively immature T cells in the early post-transplant period might expand the pool of potential alloreactive T cells and increase the risk of subsequent development of cGVHD (201).

B lymphocytes are also involved in the pathogenesis of cGVHD and both donor and recipient B cells play a role in the pathogenesis of the disease. Patients with chronic GVHD frequently have circulating antibodies that are reactive with recipient cells (190).

There are antibodies directed against host antigens which are not present in the donor, such as antibodies directed against Y-chromosome–encoded proteins that develop in male patients who receive stem cell grafts from female donors. They may predict the occurrence cGVHD and seem to correlate with disease mortality and remission (**204-206**). Antibodies directed against non-polymorphic autoantigens may be also involved in the pathogenesis of cGVHD (**153, 207**). Anti-PDGF antibodies may stimulate collagen 1 production and play a role in the development of fibrosis, as suggested in SSc (**150, 153**). In addition, it has been hypothesized that allo- and auto-antibodies may facilitate the cross-presentation of antigens to T effector cells, thereby amplifying T cell responses to minor histocompatibility antigens or autoantigens (**208**). Jin et al demonstrated that antibodies from donor B cells contributed to the damage in the thymus, spleen and lymph nodes, increased Th17 infiltration in the skin, and perpetuated sclGVHD in a murine model (**209**). B lymphocytes may contribute to the pathogenesis of cGVHD by other mechanisms such as antigen presentation and production of cytokines and chemokines (**208**). It has been shown that donor B cells in cGVHD increase pathogenic CD4⁺ T cell expansion via their antigen-presenting function (**210**). There is an impairment of normal B-cell homeostasis exists in cGVHD, similar to autoimmune diseases (**211**). High levels of BAFF in the presence of low numbers of naive B cells have been proposed to promote the survival of activated alloreactive and autoreactive B cells, leading to immune pathology (**211**). In the context of BAFF excess, activated B cells are resistant to apoptosis and display increased B cell receptor (BCR) responsiveness to antigens (**212**). In addition, the reduction regulatory B cells was also found in cGVHD patients and in murine sclGVHD models (**213, 214**).

Fibrosis is one of the distinctive features of sclGVHD. It has been hypothesized that CD4⁺ T lymphocytes release Th2 cytokines (such as IL-13, IL-10, and IL-4) which can stimulate fibroblasts resulting in sclerodermatous features (**183, 215**). Th2 cytokines may stimulate other cells such as macrophages, eosinophils, and mast cells to foster fibroblast proliferation and collagen production (**215**). Later studies have shown that IFN- γ is a prominent cytokine in the early phase of murine Scl GVHD (**216, 217**). IL-2 and TNF- α are also important for cGvHD initiation and propagation (**218**). Excessive STAT3

signaling seems to contribute to GvHD development, as prolonged STAT3 activation is prevalent in cGvHD target organs **(219, 220)**.

It seems that cytokine profiles shift over the course of fibrosis in Scl-cGVHD, changing from Th1 predominance in the early phase to a Th2 profile in the later phase **(217, 214)**. Nevertheless, the role of IL-17 e Th17 in the pathogenesis of skin fibrosis in sclGVHD have also been hypothesized **(221, 222)**.

TGF β is the critical cytokine driving fibrosis in ScGVHD. It has been related to the development of skin fibrosis in mouse models of sclGVHD **(223-225)** and in humans TGF- β levels are increased significantly during cGVHD **(226)**. On the other side, TGF- β plays an important role in the generation and maintenance of peripheral Tregs and enhancement of their suppressive function **(227)**. Anti-TGF β treatment and naturally occurring antagonist to TGF- β , human latency-associated peptide may prevent skin fibrosis in murine sclGVHD models **(228)**. In addition, Dees et al have recently demonstrated that Riociguat, a stimulator of the soluble guanylate cyclase which is able to inhibit TGF- β signalling reduced skin thickening, myofibroblast differentiation and accumulation of collagen in mouse model of sclGVHD, exerting a potent anti-fibrotic effect **(229)**.

PDGF seem to be also involved in the pathogenesis of fibrosis in sclGVHD. As previously mentioned, stimulatory antibodies to the PDGFR, present in patients with cGVHD **(153)** stimulate type I collagen production and fibroblast into myofibroblast conversion **(150, 153)**.

1.2.5. SSc versus sclGVHD: similarities and differences

Both SSc and sclGVHD are characterized by immune-mediated fibrosis where the excessive activation of T and B cells trigger and perpetuate collagen production and deposition. Indeed, the sclGVHD murine models have been used as preclinical models of SSc over the past decades (**228, 230**).

T lymphocytes play a pivotal role in the pathogenesis of both diseases and the contribution of T helper lymphocytes is essential (**97, 190**). In sclGVHD there seem to be an early polarisation and activation of Th1 lymphocytes with IFN production, while in the later phases there is a shift to predominant Th2 response and release of profibrotic cytokines, such as IL-4 and IL-14 (**231**). On the other hand, in SSc there is an early immune Th2 polarization (**142**), and later shift from Th2 to Th1 along with the disease duration, especially in patients with dSSc (reviewed in **144**). To note that the natural course of dSSc is usually characterized by spontaneous improvement or resolution of skin fibrosis (**232**).

B lymphocytes also contribute to the pathogenesis of both SSc (**149**) and sclGVHD (**208**). Anti-PDGFR antibodies are present in both pathologies (**150, 153**) and they have been shown to stimulate collagen production and miofibroblast proliferation (**233**). In addition, it has been reported that combined inhibition of c-Abl and PDGF receptors was effective for prevention and treatment of sclGVHD in mice (**233**).

TGF β is a central player of fibrotic process in both pathologies and anti-TGF β treatment seems to prevent fibrosis in sclGVHD (**223, 228, 229**) and to improve skin involvement in SSc patients (**234**).

From the clinical point of view, both pathologies are characterized by skin sclerosis. The comparative histological analysis of the fibrotic skin in SSc and sclGVHD showed similar degree of fibrosis, strong hyaluronan staining in the dermis and epidermidis, increased number of myofibroblasts and microvascular intimal hyperplasia with increased layers of smooth muscle cells (**235**).

Two interesting studies compared global genetic expression patterns of a sclGVHD murine model with transcript profiles seen in skin from SSc patients (**236, 237**). The aim

of the studies was to assess whether the phenotypic similarities between two diseases were related to the common pathogenesis or to the convergence of tissue damage patterns through separate mechanisms (**236, 237**). In SSc, the genome-wide expression profiling of fibroblast belonging to different clinical subsets of SSc patients, has identified four different profiles (fibroproliferative, inflammatory, limited, and normal-like) characterized by different expression of genes involved in diverse cellular functions, with TGF β signalling spanning the fibroproliferative and inflammatory subsets (**238-241**). In mice with sclGVHD, there was an up-regulation of genes involved in antigen presentation, immune response and protein degradation (**236**). The comparison of mice sclGVHD and human SSCc fibroblast patterns showed that sclGVHD genes clustered with the inflammatory subset of SSc (**236, 237**). The study of the gene expression data in both diseases underlined the role of IL-13 pathway activation and CCL-2 (MCP1) up-regulation in both SSc and sclGVHD (**236**). Moreover, the blockade of CCL2 prevented sclGVHD and the expression of IL-13 pathway genes and CCL2 correlated with the extent of skin fibrosis in SSc (**236**).

Although SSc and sclGVHD share clinical, histopathological and pathogenic features, there are also important differences.

The main difference consists in the fact that sclGVHD is directly related to transplant procedure, while the exact etiological agent of SSc is still unknown. Indeed, donor T cells reactive to recipient's antigens are essential for the sclGVHD development (**178**) and B donor cells are also important in disease pathogenesis (**208**).

Both diseases are characterized by loss of self tolerance, and autoantibodies are found in both. Antinuclear, anticentromere, anti-Scl70 (anti-topoisomerase I) and anti-RNA polymerase III antibodies are strongly associated with SSc (**242**). However, the role of autoantibodies in the pathogenesis of sclGVHD is uncertain. Although they are commonly found in sclGVHD, the autoantibodies positivity has typically low titer and lacks the specificity observed in primary autoimmune diseases (**171**).

In addition, the thymus involvement and the loss of central tolerance are of pivotal importance in the pathogenesis of sclGVHD (**180, 190**), whilst it seems not to be relevant

for the pathogenesis of SSc, although the thymic alterations, particularly thymic hyperplasia, have been reported in SSc patients (243).

On the other hand, endothelial damage and microvascular impairment are fundamental in the pathogenesis of SSc (97), but not of sclGVHD (235). In SSc, endothelial damage and vascular tissue insult precede and initiate fibrosis (102). Typical nailfold capillary alterations and prominent vascular manifestations, such as Raynaud's phenomenon, digital ulcers and pulmonary arterial hypertension, characterize SSc (244). In sclGVHD there are no characteristic nailfold capillary abnormalities observed in SSc (245, 246) and sclGVHD patients rarely present vasculopathy and Raynaud's phenomenon (173). However, although the skin biopsies of sclGVHD patients does not show capillary rarefaction, they present intimal hyperplasia (235).

Another histopathological difference between two diseases is the fact that in sclGVHD the fibrosis commences in the superficial skin layers and extends to deeper layers (176, 235), whereas in SSc fibrosis begins in the lower dermis and then extends upwards (235). From the clinical point of view, the main distinctive feature of SSc compared to sclGVHD is the major rate and severity of internal organ involvement (176, 244).

Nevertheless, both diseases are characterized by significant morbidity, limitation of a range of motion and impaired quality of life, related principally to the skin sclerosis.

1.3. HLA class II alleles in systemic sclerosis and sclGVHD

1.3.1. HLA class II alleles in SSc

Numerous studies have demonstrated the key role of HLA alleles in the SSc susceptibility (247-250). HLA-class II represents the most significant region associated with SSc and the major associations have been reported for HLA –DRB1, HLA-DQB1 and HLA-DPB1 alleles. Three large genome-wide association studies (GWAS) (251-253) and one immunochip study (254) in fact demonstrated that the strongest genetic associations with SSc fell within the HLA class II region. Besides the susceptibility to SSc, the studies have demonstrated correlations between certain HLA-class II alleles and each of the SSc-specific autoantibodies (250, 251, 254, 255) and between the alleles and some of the clinical manifestations of SSc, such as pulmonary fibrosis (255, 256). Furthermore, the association of the HLA class II alleles and disease mortality has been reported (257). The results vary in different ethnic populations and have been shown to vary according to the presence/absence of the three most common SSc-related autoantibodies (anti-centromere, anti-topoisomerase and anti-RNA polymerase III). However, data on the role of single HLA alleles in different populations are often contrasting.

1.3.2. SSc and HLA-DRB1 alleles

The possible role of HLA – DRB1 alleles in susceptibility and protection to SSc has been reported for different ethnical groups.

In a large multi-ethnic US study on 1300 SSc patients and 1000 controls of Caucasian, African-American and Hispanic origin, DRB1*11:04 was strongly associated with the risk of SSc, while DRB1*07:01 was protective for SSc in the US Caucasians and Hispanic individuals (250). Additionally, DRB1*01:01 was associated with increased risk of SSc in patients who were ACA positive, while DRB1*11:04 was a major risk factor for ATA positive SSc. However, these observations were different for SSc in African Americans in

whom DRB1*08:04 was found to be significantly associated with SSc in the overall group of patients and most strongly associated in ATA positive group **(250)**. The protective effect of DRB1*07:01 was confirmed in Spanish **(258)** and Han Chinese patients **(255)** and in cohort of Australian scleroderma patients, containing mainly Caucasian subject, where protective effect of DRB1*07:01 was strongest against ACA positive disease, without evidence of protection against ATA antibody positive disease **(259)**.

The association of DRB1*11 and risk of SSc was supported by studies in Spanish **(258)**, UK Caucasian patients **(260)**, Han Chinese **(255)** and in Black South African populations **(261)**. DRB1*11:04 resulted strongly associated with risk of SSc in Italian and Spanish population **(256)** and in Australian scleroderma cohort **(259)**. DRB1*11:04 was confirmed as a major risk allele for ATA positive SSc patients in UK Caucasian population **(260)** and in Mexican patients HLA-DRB*11:04 was increased in subjects with dSSc **(262)**. HLA-DRB1*11 was also associated with pulmonary fibrosis in Italian and Spanish population **(256)** and correlated with high skin scores in Canadian SSc patients of Caucasian origin **(248)**.

On the other hand, in the UK Caucasian population DRB1*08 and DRB1*04 were associated with antcentromere antibody (ACA) positivity **(260)**. The association of HLA-DRB1*08:04 and SSc reported in African Americans was not observed in black South Africans which presented association of HLA- DRB*15 and ATA positivity and correlation of DRB1*03:01 with the lcSSc **(261)**. DRB1*08:02 was associated with dSSc in Japanese **(264)** and with ATA positive SSc in US Hispanic patients **(250)**.

The association of HLA – DRB1*15:02 and ATA positive SSc patients was reported in Japanese **(264)**, Korean **(265)** and Thai patients **(266)**, whilst in Chinese Han patients HLA – DRB1*15:02 was associated with SSc ACA positive SSc **(255)**.

In addition, DRB1*15 and *16 were strongly associated with pulmonary fibrosis along with dcSSc and ATA in Chinese SSc patients **(255)**; on the contrary, in black South Africans DRB1*11 was associated with pulmonary fibrosis **(261)**.

The meta-analysis of Italian and Spanish SSc patients showed an association between HLA-DRB1*01 alleles and ACA positivity **(256)**. HLA-DRB1*01 and HLA - DRB1*04 alleles have been related to ACA positive SSc patients independently from their ethnic origin in

a small cohort of US patients of Caucasian, Hispanic and African American origin (**247**), HLA-DRB*04 was associated with ACA positive UK Caucasian patients (**260**) and DRB1*01:01 was a risk allele for ACA positive Caucasian patients in large US SSc cohort (**250**). On the other hand, HLA-DRB1*01:01 and DRB1*04:06 were strong SSc-protective alleles in Han Chinese, especially in patients who were ACA positive and had dSSc, respectively (**255**).

Increased DRB1*03:01 in lcSSc was concordant between black South Africans (**261**) and Han Chinese (**255**), whilst in a Mexican cohort of SSc patients ISSc was associated with increased HLA-DRB1*01:01 and HLA-DRB1*01:02 (**262**). Chinese Has patients showed an association of HLA - DRB1*10 with ACA positive cases, and displayed no association with HLA-DRB1*01 or DRB1*08 (**255**).

The anti-RNA polymerase III antibody positive SSc patients were explained by HLA-DRB1*04:04 in US Caucasian patients, by HLA-DRB1*11 alleles, in both Hispanic and Caucasian US subjects by HLA-DRB1*08:04 in the African American population, in which the same allele was also associated with the production of ATA (**250**).

Azzouz et al analyzed the HLA-DRB1 susceptibility alleles for SSc in a cohort of French Caucasian population and identified a common amino-acidic motif FLEDR at positions 67 – 71 of the corresponding β chain (**267**). The FLEDR motif was significantly associated with dSSc in ATA positive patients and the risk to develop SSc was higher in patients with double presence of the FLEDR motif than the single dose (**267**).

Table 1.3 shows associations of HLA-DRB1 alleles with SSc in different ethnical populations

	Caucasian USA	Hispanic USA	African USA	Spanish (ref 258)	Spanis (ref 256)	Italian	UK	Black South African	Japanese	Korean	Thai	Chinese Han	Mexican	Australian
DRB1*01:01	Y (ACA+)	Y		Y								Protective for ACA+	Y (ACA+)	
DRB1*11:04	Y	Y			Y	Y	Y (ATA+)					Y	Y (dSSc)	Y
DRB1*08:04			Y											
DRB1*08:02		Y (ATA+)							Y (dSSc)					
DRB1*03:01 in ISSc								Y				Y		
DRB1*01 in ACA+				Y		Y								
DRB1*04 in ACA+							Y							
DRB1*08 in ACA+							Y							
DRB1*11	Y	Y		Y			Y	Y				Y		
DRB1*15:02								*(DRB1*15)	Y (ATA+)	Y (ATA+)	Y (ATA+)	Y (ACA+)		
DRB1*04:06 (protective)												Y (dSSc)		
DRB1*07:01 (protective)	Y	Y		Y								Y		Y (ACA+)

Table 1.3 Associations of HLA-DRB1 alleles with SSc

Y = Yes

1.3.3. SSc and HLA-DPB1 alleles

The HLA – DPB1 alleles have been correlated with susceptibility to SSc in a number of studies on different populations.

A GWAS study in a Korean SSc cohort revealed that the specific single nucleotide polymorphisms of HLA DPB1 region conferred the strongest susceptibility to SSc (**268**). Subtyping of HLA-DPB1 showed that HLA-DPB1*13:01 and DPB1*09:01 were the most susceptible alleles to SSc, especially in ATA positive patients (**268**). The authors confirmed the association of HLA-DPB1*13:01 and ATA positive Caucasian SSc patients (but not of DPB1*09:01), by examining the 5 SNPs with the strongest association in a large cohort of US SSc subjects (**268**).

Other studies on Caucasian population (UK and USA), supported the role of DPB1*13:01, but not DPB1*09:01, as a major susceptibility allele to SSc, and especially to ATA positive SSc (**250, 260**). Only DPB1*13:01 was significantly increased in Han Chinese SSc patients (**269**) and in black South African patients with dcSSc (**261**). On the other hand, a Japanese SSc cohort supported only a weak association of DPB1*09:01 with SSc (**270**) and a study on African American and Hispanic SSc US patients failed to show any HLA-DPB1 association with SSc (**250**).

In addition, HLA-DPB1*03:01 also appeared as a strong risk allele for SSc in Han Chinese SSc patients (**269**). In this study, HLA-DPB1*13:01 and *03:01 were significantly associated with several subsets of SSc compared to the controls. In particular, increased frequencies of both alleles were strongly associated with dcSSc, ATA positive, ACA negative SSc and SSc with pulmonary fibrosis. Since the associations of DPB1*13:01 with dSSc and ATA positive SSc also were reported in US/UK Caucasian (**250, 260**) and in black South Africans (**261**), these observations may suggest potential role of these two alleles in severe cases of SSc.

Table 1.4. shows associations of HLA-DPB1 alleles with SSc in different ethnical populations

Table 1.4 - Associations of HLA-DPB1 alleles with SSc

	Caucasian USA	Hispanic USA	African USA	Caucasian UK	Black South African	Korean	Japanese	Chinese Han
DPB1*13:01	Y			Y	Y (dSSc)	Y		Y
DPB1*09:01						Y	Y	
DPB1*03:01								Y

Y = Yes

1.3.4. SSc and HLA-DQB1 alleles

The role of HLA-DQB1 alleles in susceptibility and protection to SSc has been reported in different ethnical groups of SSc patients.

HLA - DQB1*03:01 allele was associated with susceptibility to SSc in US Caucasian, Hispanic and African-American populations (250) and DQB1 alleles encoding a non-leucine residue at position 26 (DQB1 26 epi), were associated with SSc in patients with Caucasian and Hispanic genetic background in the same study (250). The association of HLA - DQB1*03:01 and disease susceptibility was confirmed in Italian and Spanish patients (256), but not in Koreans (265) and Chinese Han population (272) and only with dcSSc in black South Africans (261). Moreover, HLA - DQB1*03:01 was correlated with ATA in USA Caucasian SSc patients (250). However recent data on Mexican patients with Caucasian genetic background showed that DQB1*03:01 allele was more frequent in healthy controls than in SSc patients, suggesting a possible protective role of this allele in this population (262)

HLA- DQB1*06:11 was increased in Han Chinese patients with SSc and more commonly in those which were ATA positive (272).

HLA- DQB1*02:02 was negatively correlated and possibly protective in dominant model in Caucasian and Hispanic US patients (250) and HLA- DQB1*06:02 was possibly protective only in US Caucasian patients (250) in the same study. To note that HLA - DQB1*03:01 is in linkage disequilibrium with DRB1*11 alleles and that HLA - DQB1*06:02 is in linkage disequilibrium with HLA-DRB1*15:01 alleles (250, 267).

In a large GWAS study that included more than 5000 SSc patients of European ancestry, HLA-DQB1 locus was correlated with ACA antibodies (249). HLA - DQB1*05:01 allele was significantly increased in US Caucasian, Spanish and Japanese patients with ACA (250, Simeon 2009, Kuwana M 1995), but not in Hispanic and African-Americans (250). In addition, DQB1*05:01 was significantly over-expressed in SSc, and was strongly associated with ACA autoantibodies in Chinese Han patients (272), as HLA – DRB1*15:02 (255). In fact, DRB1*15:02/DQB1*05:01 is considered as a haplotype in Asian population (274).

It is interesting to note that in a large cohort of Italian and Spanish patients, HLA-DQB1*05:01 allele displayed the protective role in ATA positive patients (256). Moreover, a reduced risk of progression from early to definite SSc has been recently correlated with the HLA DQ5-DR1 haplotype (HLA-DRB1*01:01/HLA-DQA1*01:01/HLADQB1*05:01) in a cohort of Italian patients (275).

HLA-DQB1 alleles that encode a non-leucine residue at position 26 (HLA-DQB1 26 epi) alleles resulted over-expressed in US and UK Caucasian and Chinese Han ACA positive SSc patients (250, 260, 272), but were also associated with ATA in African American patients (250). HLA DQB1*06:01 haplotypes were reported in Japanese patients with dcSSc and ATA (264). In the Mexican population of Amerindian origin, a significant association for the presence of ATA antibody with HLA-DQB1*04:02 has been observed (262).

The anti-RNA polymerase III antibody (ARA) positive SSc patient subset was explained by HLA-DQB1*03 in Caucasian and Hispanic populations (250), but by HLA-DRB1*08 in black African Americans (250). No association between HLA - DQB1 and SSc was found in the Australian SSc cohort composed mostly of Caucasian patients (259).

HLA-DQB1 susceptibility alleles for SSc, DQB1*03:01, *04:02 and *06:01 code for a common TRAE_{LD}T^{motif} at positions 71-77 of their β chain. The addition of a tyrosine at

the position 30 strengthened the TRAE LDT association with dSSc (**247, 267**). TRAE LDT was associated with dSSc and ATA positivity in French Caucasian patients with a higher risk for in case of double presence of the shared epitope (**267**).

Table 1.5 shows associations of HLA-DQB1 alleles with SSc in different ethnical populations

	Caucasian USA	Hispanic USA	African USA	Spanish (ref 258)	Spanish (ref 256)	Italian	UK	Black South African	Japanese	Chinese Han	Mexican
DQB1*03:01	Y	Y	Y	Y	Y	Y		Y (dSSc)			Protective
DQB1*02:02 protective	Y	Y									
DQB1*05:01 ACA +	Y			Y	Protective for ATA+	Protective for ATA +			Y	Y	
DQB1*04:02											Y (ATA+)
DQB1*06:01									Y (ATA+)		
DQB1*06:02 protective	Y										
DQB1*06:11										Y	
DQB1 26 epi	Y	Y					Y (ACA+)			Y (ACA +)	

Table 1.5 - Associations of HLA-DQB1 alleles with SSc

Y = Yes

1.3.5. SSc and HLA-DQA1 alleles

Several studies have shown the association of HLA-DQA alleles and SSc. The association of HLA-DQA1*05:01 and SSc have been reported in Caucasian and Hispanic US patients (250), as well as in Italian and Spanish (256) and in Choctaw Amerindians (277, 278). HLA-DQA1*02:01 seemed to have a protective role in Caucasian and Hispanic US patients (250).

Mayes et al. performed a large study on 5850 SSc patients and 9401 controls of European origin, genotyped with the ImmunoChip array (254). Classical alleles, amino acid residues, and SNPs across the HLA region were imputed and tested. The authors analysed the associations of the following HLA class II four digit alleles previously reported to be correlated with SSc or with its clinical subsets: HLA-DQB1*05:01, HLA-DPB1*13:01, HLA-DQA1*02:01, HLA-DQA1*05:01, HLA-DQB1*02:02, HLA-DQB1*03:01, HLA-DRB1*11:04, HLA-DRB1*07:01 and HLA-DRB1*15:01. The association with SSc was confirmed for all the alleles ad exception of DLA-DRB1*07:01 (254).

1.3.6. SclGVHD and HLA class II alleles

There is a number of studies which investigated a role of donor-recipient HLA class II alleles mismatch in the pathogenesis of sclGVHD (171, 174, 175). On the other hand, there is only one report on the role of HLA class II alleles in the genetic susceptibility towards sclGVHD (279). Inamoto et al recently reported that the risk of sclGVHD was decreased in patients who carried HLA-DPB1 alleles which codified for HLA-DP molecules with glutamine (Q) at position 31 of the α chain and GPM (glycine-proline-methionine) motif at positions 85-87 of the β chain (279). These amino acidic positions are located within the P1 peptide-presentation pocket of HLA-DP molecule.

2. AIM

SSc is an autoimmune chronic disease characterized by excessive deposition of collagen and other components of extracellular matrix in skin and internal organs, prominent and often severe microvascular alterations and humoral and cellular immunologic abnormalities (17). The fibrosis of skin and internal organs is one of the main causes of tissue damage, and the severity and rate of progression of tissue fibrosis is of paramount importance in determining prognosis in SSc. Indeed, cutaneous fibrosis correlates with both survival and functional limitations in SSc (99, 280). To date there are no predictive markers of severity of skin or other organ fibrotic involvement in SSc.

CGVHD is the most common non-relapse complication affecting long-term survivors of HSCT (168, 169). cGVHD skin involvement during cGVHD is heterogeneous and poorly predictable among patients. The sclerotic variant of cGVHD (sclGVHD) shares many similarities to SSc skin involvement and often is poorly responsive to therapy (171). Scl-cGVHD is associated with high morbidity, limited functional mobility and decreased quality of life (171,173). Likewise in SSc, predictive markers of occurrence or severity of sclerotic skin involvement in cGVHD are lacking.

Given the clear immunologic basis of tissue damage both in scl-GVHD and SSc, genetic factors involved in the immune response may contribute to the heterogeneous nature of tissue fibrosis in both conditions. Indeed, genome wide association studies (GWAS) conducted in SSc have identified the HLA class II region as most strongly associated with SSc (252). Nevertheless, given the extremely high polymorphism of HLA, GWAS are limited to screening for single nucleotide polymorphisms in non-hypervariable regions and do not consider the complementarity of the determining regions for the antigen loading. The classic HLA association studies proposed a number of different class II alleles as risk factors for SSc or for its clinical manifestations (250, 256, 258). Unfortunately, patients with different disease presentations and internal organ involvement are commonly included in SSc studies, hampering the validity of association studies aimed to identify precise risk factors. The clinical heterogeneity of SSc, the low

prevalence of the condition, and the apparently inconsistent results of the above mentioned genetic studies have, therefore, mitigated interest in specific HLA haplotypes as potential risk factors in the immunopathogenesis of tissue fibrosis.

More than two decades ago a similar scenario was dominating the genetic studies aimed at HLA associations with Rheumatoid Arthritis (RA). Strikingly, in the era of serologic HLA typing, a meta-analysis study observed that HLA-DRB1 alleles reported to be associated with RA shared the RAA (arginine, alanine and alanine) amino acid motif at positions 72–74 of the third hypervariable region, which they hypothesized to act as a functional unit and called shared epitope (SE) (50). Since that discovery, a multitude of studies have confirmed and extended that association, providing a foundation for understanding the genetic contribution to RA clinical heterogeneity. It is now well established that HLA-DRB1 SE alleles are associated with susceptibility to RA and severity of the disease (64, 65, 281).

The analysis of the HLA hypervariable regions in SSc and sclGVHD patients may reveal the existence of common functional motifs (shared epitopes), similar to those demonstrated in RA. These motifs could be of use both as predictive factors of severity in scl-GVHD or SSc and might shed light into the molecular mechanisms underlying the still unclear immunopathogenesis of tissue fibrosis.

The aim of this study was to identify the HLA class II functional motifs (shared epitopes) associated with severe skin fibrosis in SSc and scl-GVHD.

In order to obtain shared functional motifs we:

- Performed high resolution four-digit genotyping of a cohort of SSc patients with heterogeneous degree of skin fibrosis
- Identified sclGVHD patients with severe skin involvement from a cohort of cGVHD patients enrolled in the in the National Cancer Institute (USA) observational clinical trial on natural history of the disease and performed high resolution four-digit genotyping in patients who did not have it available

- Obtained and analyzed amino acidic sequences corresponding to the hypervariable regions of β chains of HLA class II molecules (DPB1, DQB1 and DRB1) in SSc and sclGVHD patients
- Compared these amino acidic sequences of sclGVHD and SSc patients with severe skin fibrosis (fibrotic phenotype) with data of SSc patients with prevalent vascular involvement and minimal skin fibrosis (non-fibrotic phenotype)
- Performed quantitative analysis of the similarity of hydrogen bonds strength within the different populations to determine whether beyond identity, functionally similar motifs could be associated with severe fibrosis.

3. MATERIALS AND METHODS

3.1. Patients

The study population consisted of:

- 45 patients with sclGVHD and clinical evidence of skin fibrosis involving more than 50% of body surface area, with histological confirmation of the fibrotic involvement on skin punch biopsy.
- 70 patients with SSc: 15 patients with diffuse cutaneous subset (dSSc) and 55 controls with limited cutaneous subset (lSSc).

3.1.1. SclGVHD patients selection and evaluation

Patients were identified from a cohort patients enrolled in the National Cancer Institute (NCI), USA, clinical trial named: “Natural History Study of Clinical and Biological Factors Determining Outcomes in Chronic Graft-Versus-Host Disease” (NCI study 04-C-0281) (282). At the time of the patients’ enrolment in our study, the NCI study 04-C-0281 cohort included more than two hundred prospectively enrolled cGVHD patients (2/3rd of them with severe NIH stage cGVHD), exceptionally well annotated for their clinical characteristics. Patients with clinical evidence of fibrotic changes (morphea-like, SSc-like, and fasciitis-like) with greater than 50% body surface area were included in the study.

All evaluations were collected during a single visit to the NIH in Bethesda by the use of predefined clinical forms as per NCI study 04-C-0281 study protocol. The evaluation was performed by a multi-specialist team. Demographic and clinical data, as well as laboratory, imaging and histopathologic features were collected. Skin involvement was assessed by a dermatologist expert in cGVHD. The extent of skin involvement was

expressed by percentage of body surface area (BSA) according to the rule of nines (Figure 3.1).

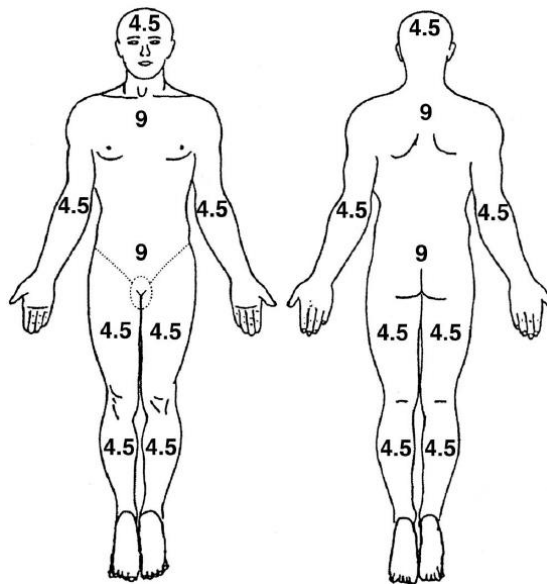


Figure 3.1. - Evaluation of the body surface involvement.

Clinical evidence of sclGVHD was determined by the presence of skin thickening, nodularity or rippling of subcutaneous tissues on deep palpation. BSA involvement was recorded for the presence of dermal sclerosis (superficial sclerosis) and subcutaneous tissue sclerosis (deep sclerosis). Total sclerosis comprising both superficial and deep sclerosis was also recorded. Histological confirmation was obtained by punch skin biopsy in all patients.

NIH cGVHD organ-specific scores were determined according to the NIH consensus criteria (170). NIH global scores were defined as “mild” (1-2 organs with scores 1), “moderate” (> 2 organs with score 1, any score 2, or lung score 1), or “severe” (any score of 3 or lung score 2), as proposed by Filipovich et al (170).

3.1.2. SSc patients selection and evaluation

Eighty consecutive SSc patients referring to one single centre in northern Italy (IRCCS Ca' Granda - Ospedale Maggiore Policlinico di Milano) were enrolled. Ten of them had

ambiguous alleles and unclear genotyping and were excluded from the study. All patients fulfilled 2013 ACR/EULAR classification criteria (283) and were categorized as having the ISSc or dSSc subset of the disease, according to LeRoy (91). Patients were periodically evaluated with complete clinical examination, laboratory tests, pulmonary function testing and echocardiography; the suspicion of pulmonary arterial hypertension and interstitial lung disease was confirmed by right heart catheterization and high-resolution chest tomography, respectively. The patients' demographic, historical, clinical, imaging and laboratory features, including the autoantibody profile were obtained by reviewing the patients' medical records.

Skin involvement was determined by the modified Rodnan skin score (mRSS) with assessment of 17 anatomical sites (face, fingers, hands, forearms, upper arms, chest, abdomen, thighs, legs, and feet) and scoring on a 0–3 scale, where 0 was normal skin, 1 was slight thickening, 2 was moderate thickening, and 3 was hidebound skin sclerosis. The scores for all sites were summed to give a total score, with a maximum possible score of 51 (284). The mRSS was assessed by a physician trained and experienced in the skin scoring technique.

The study was approved by local ethical committees in Milan (U.O. Comitato di Etica e Sperimentazione Farmaci Fondazione IRCCS Ca' Granda - Ospedale Maggiore Policlinico di Milano) and in Leeds (Regional ethical committee of Yorkshire and the Humber) and by the Institutional review board of the National Cancer Institute, USA.

3.2. HLA typing

High resolution HLA typing of SSc patients and of sclGVHD patients without 4-digit HLA typing was performed by commercially available kits (Fujirebio Diagnostics, Inc):

- INNO-LiPA HLA-DPB
- INNO-LiPA HLA-DRB1 Plus
- INNO-LiPA HLA-DQB1 Update

LiPA are line probe assays based on the polymerase chain reaction (PCR) followed by reverse dot-blot hybridization of the PCR products. DNA was extracted from the peripheral blood and the genotyping was performed according to the manufacturer instructions. Amplified, biotinylated DNA material is chemically denatured, and the single strands are hybridized with specific oligonucleotide probes immobilized as parallel lines on membrane-based nitrocellulose strips. This process is followed by a stringent wash step to remove any mismatched amplified material. Thereafter, streptavidin conjugated with alkaline phosphatase is added and bound to any biotinylated hybrid previously formed. Incubation with a substrate solution containing a chromogen results in a purple-brown precipitate. The reaction is stopped by a wash step, and the reactivity pattern of the probes is recorded.

3.3. HLA data alignment and analysis

HLA amino-acidic sequences were obtained from complete 4 digit typing data using a web based HLA alignment tool of the EMBL-EBI (31).

The amino acids derived from the hypervariable regions of each HLA allele were extracted and recorded in a specific database. The amino acids corresponding to the hypervariable regions were located in the following positions:

- HLA-DPB1: 8-11, 35-36, 55-57, 65-69 and 84-88
- HLA-DQB1: 26-37, 52-57, 70-74
- HLA-DRB1: 9-13, 25-28, 67-74

The recognition of shared motifs was conducted by identification of identical residues or functionally similar residues.

For the identification of functionally similar residues two parallel approaches were applied:

- a) Qualitative: the amino acids were classified according the three classic categories (non-polar; polar-neutral; polar-charged) and two subcategories(Hydrophobicity/Hydrophilicity).
- b) Quantitative: according to Rogov et al. (285) in which the aminoacids were classified according to a numerical measure function of the number of peptide bonds able to harbour in their surroundings.

3.3.1. Qualitative analysis

The classification of amino acids according to their physical, chemical and structural properties proposed by Taylor et al (286) was applied (Figure 3.2).

The first classification of amino acids into categories (families) used for the purpose of the study was the following:

- Polar + small/very small: N, S, C, T
- Polar + charged negative + small: D
- Polar + charged negative: E
- Polar + charged positive: R, K, H
- Polar only: W, Y, Q
- Non polar only: F, M, I, L
- Non polar + small/very small: V, A, G, P

Since the most important features for the epitope binding are represented by polarity and hydrophobicity, we applied also the second classification of amino acids in:

- Hydrophobic (M, I, V, L, H, Y, W, F, K, C, T, A) and Not-hydrophobic (G, S, D, E, R, P, N, Q)
- Polar (Q, N, S, C, T, Y, W, H, K, R, D, E) and Not-polar (P, G, A, I, V, L, M, F). Polar amino acids were divided in positive charged (R, K, H), negative charged (D, E) and not-charged (Q, N, S, C, T, Y, W).

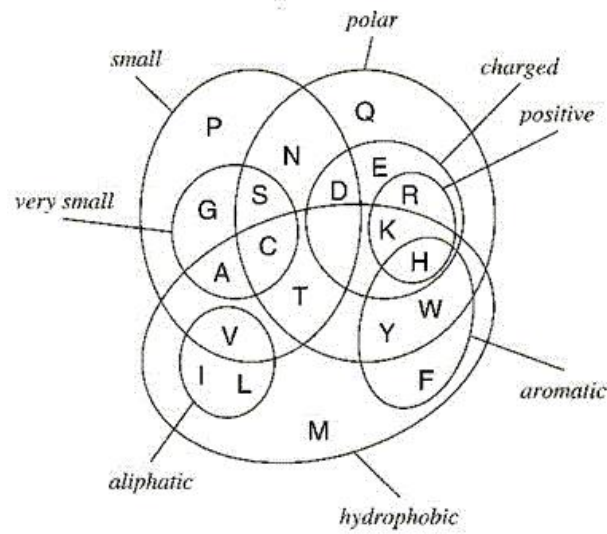


Figure 3.2. - Classification of amino acids according to Taylor et al. (adapted from reference 286).

Amino acids abbreviations:

- W = Tryptophan
- Y = Tyrosine
- Q = Glutamine
- N = Asparagine
- S = Serine
- C = Cysteine
- T = Threonine
- D = Aspartic acid
- E = Glutamic acid
- R = Arginine
- K = Lysine
- H = Histidine
- F = Phenylalanine
- M = Methionine
- I = Isoleucine
- L = Leucine
- V = Valine
- A = Alanine
- G = Glycine
- P = Proline

3.3.2. Quantitative analysis

For the quantitative analysis of amino acids, we utilized the coefficients of similarity of amino acids, named measure of similarity of environments of residues (m) proposed by Rogov et al (285) (Table 3.1).

Rogov elaborated a measure of similarity of surroundings of the amino acid residues in primary structures of native proteins, which is a continuous numerical criterion and is based on the number of peptide bonds (from 1 to 20) that each amino acid was able to form with other amino acids in its surrounding. M are nearer to 1 for those residues which surroundings display a higher degree of similarity. The increase in dissimilarity of the residues surroundings corresponds to a decrease in m .

	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A	1.000	-0.168	0.029	-0.033	0.064	-0.023	0.014	0.068	0.031	0.140	-0.562	0.131	0.020	0.051	0.143	0.249	0.248	0.232	-0.234	0.011
C	-0.168	1.000	-0.406	-0.672	-0.082	-0.257	-0.220	-0.159	-0.527	-0.085	-0.787	-0.298	-0.298	-0.449	-0.392	-0.147	-0.156	-0.091	-0.464	-0.199
D	0.029	-0.406	1.000	0.109	-0.227	-0.034	-0.003	-0.336	0.036	-0.314	-0.839	0.329	0.005	0.0676	0.026	0.215	0.105	-0.182	-0.446	-0.203
E	-0.033	-0.672	0.109	1.000	-0.432	-0.299	-0.284	-0.465	0.224	-0.432	-0.877	-0.010	-0.218	0.183	0.090	-0.064	-0.145	-0.317	-0.461	-0.352
F	0.064	-0.082	-0.227	-0.432	1.000	-0.097	-0.016	0.380	-0.277	0.367	-0.529	-0.073	-0.067	-0.217	-0.113	0.143	0.148	0.345	0.021	0.352
G	-0.023	-0.257	-0.034	-0.299	-0.097	1.000	-0.118	-0.256	-0.239	-0.205	-0.806	0.044	0.044	-0.176	-0.119	0.120	-0.039	-0.080	-0.375	-0.178
H	0.014	-0.220	-0.003	-0.284	-0.016	-0.118	1.000	-0.130	-0.111	-0.087	-0.785	0.167	-0.031	0.0252	0.084	0.148	0.103	-0.055	-0.262	0.073
I	0.068	-0.159	-0.336	-0.465	0.380	-0.256	-0.130	1.000	-0.250	0.486	-0.452	-0.124	-0.187	-0.259	-0.139	-0.026	0.129	0.494	-0.021	0.237
K	0.031	-0.527	0.036	0.224	-0.277	-0.239	-0.111	-0.250	1.000	-0.236	-0.712	0.197	-0.171	0.287	0.387	0.028	-0.010	-0.167	-0.364	-0.203
L	0.140	-0.085	-0.314	-0.432	0.367	-0.205	-0.087	0.486	-0.236	1.000	-0.389	-0.123	-0.159	-0.180	-0.090	0.015	0.092	0.365	0.002	0.189
M	-0.562	-0.787	-0.839	-0.877	-0.529	-0.806	-0.785	-0.452	-0.712	-0.389	1.000	-0.682	-0.682	-0.742	-0.681	-0.610	-0.548	-0.514	-0.717	-0.644
N	0.132	-0.298	0.329	-0.010	-0.073	0.044	0.167	-0.124	0.197	-0.123	-0.682	1.000	0.063	0.161	0.183	0.271	0.241	-0.055	-0.309	-0.045
P	0.020	-0.298	0.005	-0.218	-0.067	0.044	-0.031	-0.187	-0.171	-0.159	-0.767	0.063	1.000	-0.083	-0.052	0.194	0.140	-0.034	-0.354	-0.119
Q	0.051	-0.449	0.068	0.183	-0.217	-0.176	0.025	-0.259	0.287	-0.180	-0.742	0.161	-0.083	1.000	0.375	0.085	-0.002	-0.143	-0.251	-0.112
R	0.143	-0.392	0.026	0.090	-0.113	-0.119	0.084	-0.139	0.387	-0.090	-0.681	0.183	-0.052	0.375	1.000	0.127	0.062	-0.037	-0.168	0.003
S	0.249	-0.147	0.215	-0.064	0.143	0.120	0.148	-0.026	0.028	0.015	-0.610	0.271	0.194	0.085	0.127	1.000	0.455	0.129	-0.249	0.052
T	0.248	-0.156	0.105	-0.145	0.148	-0.039	0.103	0.129	-0.010	0.092	-0.548	0.241	0.140	-0.002	0.062	0.455	1.000	0.247	-0.209	0.087
V	0.232	-0.091	-0.182	-0.317	0.345	-0.080	-0.055	0.494	-0.167	0.365	-0.514	-0.055	-0.034	-0.143	-0.037	0.129	0.247	1.000	-0.015	0.246
W	-0.234	-0.464	-0.446	-0.461	0.021	-0.375	-0.262	-0.021	-0.364	0.002	-0.717	-0.309	-0.354	-0.251	-0.168	-0.249	-0.209	-0.015	1.000	0.080
Y	0.011	-0.199	-0.203	-0.352	0.352	-0.178	0.073	0.237	-0.203	0.189	-0.644	-0.045	-0.119	-0.112	0.003	0.052	0.087	0.246	0.080	1.000

Table 3.1 - M values proposed by Rogov et al.
(Adapted from reference 285)

For the quantitative analysis we identified the sequences expressed differently between the two groups of patients in a significant manner and selected the longest one (one for every of the three HLA genes) as a reference sequence.

The reference sequence is formed by amino acids at determinant positions, derived from the same allele (maternal or paternal). Since every patient had two possible sequences for every series of amino acids at determinant positions, we assessed Rogov coefficients for both amino acids of the copy, comparing them to the reference amino acid at these position. In order to normalize by positive or negative, we then subtracted each coefficient from 1. All the values (one for every position) for each of two amino acidic sequences were added to form two scores of diversity with the reference sequence for each patient. We then grouped the patients by the number of differences from the reference amino acid at every position. By doing so, the patients that had one reference sequence, codified by one of the alleles, were considered as not different from the reference (we presumed a dominant effect of the sequence in the phenotype expression).

The lowest score was obtained more similar was the given sequence to the reference sequence.

3.3.3. Statistical analysis

For the analysis the statistical program SPSS (version 22) and SAS/STAT software for statistical analysis (version 14.1) were used.

The evidence of independent assortment in the frequencies of the amino acids and the consistency with Hardy-Weinberg proportions were tested with chi-square based method available in the statistical softwares. The Fisher-Freeman-Halton test was used to assess the association between the amino-acids and diagnostic groups and between the amino acids and the families/categories. The multiple comparison corrections were performed with Hochberg correction and False Discovery Rate (Benjamini-Hochberg) correction. Hochberg correction, a step-up version of Bonferroni's test was performed

to control type I error rate and to reduce the risk of false positives. However, since this method may lead in turn to high number of false negatives for a large number of comparisons, we also performed alternative multiple comparison correction with less sensitive, Benjamini-Hochberg or False Discovery Rate (FDR) correction. FDR controls the expected proportion of "discoveries" (significant results) that are in reality false positives.

The tests based on the ranks of the ordered values (Spearman Rank-Order Correlation and Kendall's test Tau-b Correlation Coefficient) were applied to test the diversity between the diagnostic groups in the number of differences from the reference amino acids. The same tests were used to compare the groups if diagnosis according to the similarity score based on the Rogov's method.

Logistic regression models were used to assess the correlation between the pairs, triplets or combination of amino acids and risk of developing fibrosis. For each of these models the log-likelihood improvement in risk prediction (and its significance) over the null model was calculated.

4. RESULTS

Genotyping results for HLA-DRB1, HLA-DPB1 and HLA-DQB1 were obtained and respective amino acids located in the hypervariable regions of the examined genes were analysed. Patients with ambiguous alleles and unclear genotyping were excluded from the analysis.

The amino acid sequences were available for the following number of patients (Table 4.1):

- 43 sclGVHD patients with HLA - DPB1 results, 40 sclGVHD patients with HLA- DQB1 results and 41 sclGVHD patients with HLA- DRB1 results.
- 70 systemic sclerosis (SSc) patients with HLA - DPB1, HLA - DQB1 and HLA - DRB1 results. There were 15 subjects with diffuse cutaneous subset (D) and 55 with limited cutaneous subset (L)).

Table 4.1 Study population

	HLA – DPB1	HLA – DQB1	HLA – DRB1
sclGVHD	43	40	41
SSc	70	70	70

Clinical characteristics of the study population are shown in Table 4.2 (features of sclGVHD patients) and Table 4.3 (features of SSc patients). Patients with ISSc had the mean mRSS of $6,3 \pm 3,4$. Skin fibrosis reached the mean of $67,4 \pm 23,2\%$ of a total body surface area in sclGVH patients.

Table 4.2 - Clinical features of sclGVHD patients (43 patients)

Age at onset of sclGVHD (years)	44,9 ± 14
Gender no. (%)	
Female	18 (41,9%)
Male	25 (58,1%)
Patient race, no. (%)	
Caucasian	35 (81,3%)
African American	4 (9,3%)
Asian	2 (4,7%)
Other	2 (4,7%)
Disease that led to HSCT	
ALL, AML, MDS	28 (65,2%)
CML, IMF, and myeloproliferative disorders	6 (13,9%)
CLL	3 (6,9%)
HL, NHL	3 (6,9%)
MM	2 (4,8%)
AA+PNH	1 (2,3%)
Stem cell graft source, no (%)	
Bone marrow	11 (25,6%)
Mobilized blood cells	32 (74,4%)
HLA and donor type, no (%)	
HLA matched related	19 (44,2%)
HLA matched unrelated	21 (55,8%)
Interval between HSCT and sclGVHD onset (months)	16,9 ± 38,7
Autoantibodies positivity	
ANA	13 (30,2%)
ENA	2 (4,8%)
ACA	0
ATA	0

Dermal fibrosis (%)	24,6 ± 25,34
Deep fibrosis (%)	37,64 ± 24,50
Total skin fibrosis (%)	67,44 ± 23,22
Organ involvement %	
Eye	12 (27,9%)
Mouth	43 (76,7%)
Gastrointestinal tract	16 (37,2%)
Liver	11 (25,66%)
Lung (bronchiolitis obliterans)	1 (2,3%)
Genital tract	2 (4,7%)
NIH stage	
Moderate	14 (33,2%)
Severe	29 (67,4%)

no	=	number
ALL	=	acute lymphocytic leukemia
AML	=	acute myeloid leukaemia
MDS	=	myelodysplastic syndrome
CML	=	chronic myeloid leukemia
IMF	=	idiopathic myelofibrosis
CLL	=	chronic lymphocytic leukemia
HL	=	Hodgkin's lymphoma
NHL	=	non-Hodgkin's lymphoma (NHL)
MM	=	multiple myeloma
AA+PNH	=	aplastic anemia + paroxysmal nocturnal haemoglobinuria
ANA	=	antinuclear antibodies
ENA	=	extractable nuclear antigen antibodies
ACA	=	anticentromere antibodies
ATA	=	anti-topoisomerase antibodies
NIH	=	National Institutes of Health

Table 4.3 Clinical features of SSc patients

Age (years)	46,2 ± 13,1
Gender	
Female	66 (94,3%)
Male	4 (5,7 %)
Subset	
Limited	55 (78,6%)
Diffuse	15 (21,4%)
Raynaud phenomenon duration (years)	11,3 ± 7,1
Disease duration (years)	8,1 ± 5,6
Autoantibodies	
ANA	66 (94,3%)
ACA	45 (64,3%)
ATA	15 (21,4%)
Videocapillaroscopy	
Early	7 (16,3%)
Active	17 (39,5%)
Late	19 (44,2%)
mRSS	10,7 ± 19,8
dSSc	26,9 ± 12
ISSc	6,3 ± 3,4
Digital ulcers	28 (40 %)
Interstitial lung disease	20 (28,6%)
Pulmonary arterial hypertension	4 (5,7%)
History of scleroderma renal crisis	2 (2,9%)

ANA = antinuclear antibodies

ACA = anticentromere antibodies

ATA = anti-topoisomerase antibodies

mRSS = modified Rodnan skin score

dSSc = diffuse systemic sclerosis

ISSc = limited systemic sclerosis

Diffuse SSc patients were divided into severe diffuse SSc (D1), defined as patients who reached the mRSS peak of at least 25 in the course of their lives (9 patients) and mild diffuse SSc (D2), defined as patients with maximum mRSS < 25 (6 patients).

Subsequently, patients were divided into two groups according to the degree of skin fibrosis:

- group 1: sclGVHD + D1 patients
- group 2: L patients

SclGVHD and D1 patients were analysed together as they were characterized by a comparable severe degree of immune-mediated cutaneous fibrosis (fibrotic phenotype). Group L was considered as a control group, being characterized principally by vascular manifestations of SSc and by low degree of cutaneous fibrosis (non-fibrotic phenotype).

The HLA class II alleles observed most frequently in sclGVH population are reported in Table 4.4.

Table 4.4 HLA class II alleles in sclGVHD

HLA – DRB1		HLA – DPB1		HLA – DQB1	
Allele	Frequency	Allele	Frequency	Allele	Frequency
DRB1*03:01	13,4%	DPB1*04:01	43%	DQB1*03:01	22,5%
DRB1*07:01	12,2%	DPB1*02:01	12,8%	DQB1*02:01	13,8%
DRB1*15:01	12,2%	DPB1*01:01	11,6%	DQB1*06:02	12,5%
DRB1*04:01	8,5%	DPB1*03:01	5,8%	DQB1*02:02	8,8%
DRB1*11:01	7,3%	DPB1*05:01	5,8%	DQB1*03:02	8,8%
DRB1*13:02	7,3%			DQB1*05:01	8,8%
DRB1*13:01	6,2%			DQB1*06:04	6,3%

The HLA class II alleles observed most frequently in SSc population are reported in Table 4.5

Table 4.5 HLA class II alleles in SSc

	Allele	SSc (frequency)	ISSc (frequency)	dSSc (frequency)
HLA - DRB1	DRB1*11:04	25%	22,7%	33,3%
	DRB1*07:01	12,1%	11,8%	13,3%
	DRB1*01:01	8,6%	9,1%	6,7%
	DRB1*08:01	6,4%	7,3%	3,3%
	DRB1*03:01	5,7%	5,5%	6,7%
	DRB1*16:01	5,7%	5,5%	6,7%
	DRB1*11:01	5%	5,5%	3,3%
	DRB1*13:02	5%	5,5%	3,3%
	DRB1*14:01	5%	4,5%	6,7%
HLA – DBP1	DPB1*04:01	25%	20%	26,4%
	DPB1*02:01	14,3%	10%	15,5%
	DPB1*04:02	13,6%	23,3%	10,9%
	DPB1*13:01	12,1%	23,3%	9,1%
	DPB1*03:01	11,4%	6,7%	12,7%
	DPB1*10:01	7%	6,7%	7,3%
HLA- DQB1	DQB1*03:01	31,4%	28,2%	43,3%
	DQB1*02:01	15,7%	16,4%	13,3%
	DQB1*05:01	13,6%	15,5%	6,7%
	DQB1*04:02	7,1%	8,2%	3,3%

	DQB1*05:03	7,1%	7,3%	6,7%
	DQB1*03:03	5,7%	4,5%	10%
	DQB1*05:02	5,7%	5,5%	6,7%

The hypervariable region of the gene DPB1 has 19 positions with respective amino acids; the DQB1 gene has 25 and the DRB1 gene has 17 positions in the hypervariable region, for a total of 61 positions. Twenty positions out of 61, had the same amino acid in all patients at both chromosomes, thus 41 positions were included in the analyses: 14 in DPB1, 14 in DQB1 and 13 in DRB1.

Before performing the comparative analysis between the groups, the presence of potential non-independent assortment in the frequencies of the amino acids was assessed. All amino acid frequencies distribution was consistent with Hardy-Weinberg equilibrium.

4.1. Comparison of amino-acidic frequencies between the groups

In the first part of the analysis we compared the frequencies of single amino acids at each position of each gene between the group 1 (fibrotic phenotype) and group 2 (non-fibrotic phenotype). Using a strict Hochberg multiple comparisons correction for the 41 tests performed, **HLA - DRB1 position 71 and HLA - DPB1 position 11** had significantly different distributions of the amino acid frequencies ($p=0.0061$ and $p=0.019$, respectively). There were four amino acids at the position 71 of HLA - DRB1 (A, E, K and R), and the biggest difference was observed between the 12% frequency of A in the group 1 and the 3,4% frequency of E in group 2. HLA - DPB1 position 11 had only amino acids G and L, with L frequencies of 20,2% in the group 1 and 40,9% in the group 2.

Using a less strict False Discovery Rate (FDR) correction, there were differences of amino-acidic frequencies between groups at **HLA-DQB1 position 26 and HLA - DRB1 position 67** ($p = 0,0018$ and $p = 0,008$ respectively). HLA - DQB1 position 26 displayed

amino acids G, L and Y; the frequencies of G were 17,3% in the group 1 and 36,4% in the group 2. There were three amino acids at position 67 on HLA – DRB1. The group 1 had amino acid F frequency of 23% and amino acid I frequency of 43%, and the group 2 had amino acid F frequency of 44,5% and amino acid I frequency of 24,5%.

In addition, the frequencies of couples of amino acids at each position (derived from maternal and paternal alleles) were analysed. After Hochberg correction of the 41 p values, the couples at **HLA - DQB1 position 26** displayed a significant difference between the groups 1 and 2 ($p=0.024$). Specifically, the LL couple was present in 33% of patients in the group 1 and only in 4% of patients in the group 2.

In the second part of the analysis we investigated whether there was any combination of amino acids associated with risk of developing fibrotic phenotype characterized by severe immune-mediated fibrosis. For this purpose we tested all possible pairs and triplets of amino acids for association with group 1 (sclGVHD + D1) or with group 2 (L). We used these amino-acidic combinations to predict risk of developing/not developing fibrotic phenotype, and calculated for each of these models the log-likelihood improvement in risk prediction (and its significance) over the null model.

4.2. Association of pairs of amino-acids with risk of developing fibrotic phenotype

a) HLA – DPB1

In the 107 patients with available amino-acidic sequence for HLA - DPB1, 535 combinations from all the pairs of amino acids at different positions were possible. Some of the possible combinations did not appear in any of the patients, and thus only 513 combinations were tested. Because of the very tight linkages between some of the positions, many of the combinations were correlated or even redundant. For this reason, only the positions that were associated with the study groups were investigated.

There was a **subgroup of 56 patients with amino acid D at position 55, E at position 56, and D, E, A and V at positions 84 through 87**, respectively, all belonging to the same chromosome (maternal or paternal). When the analysis per diagnostic group was performed, 37 of these 56 patients (66%) belonged to the group 2 (had non-fibrotic phenotype) while only 18 of the other 51 patients who did not present this DEDEAV amino acidic sequence (35%) were in the group 2. This difference was statistically significant ($p=0.0019$). However, due to the linkages, it was impossible to determine whether any outcome in these 56 patients was attributable to D at 55, E at 56 or both together in combination with any of the last four separately or together.

The amino acids at DPB1 position 36 were linked to amino acids at positions 55 and 56, although not as strongly as the ones above. The **combination of amino acid V at position 8 with V at 36 or D at 55 (equivalent to E at 56) characterized 54 of the 56 patients who presented DEDEAV sequence at positions 55,56, 84-87**. Thirty-six out of these 54 patients (67%) had non-fibrotic phenotype (group 2) versus 19 of other 53 patients (36%) who did not present this combination ($p=0.0019$).

The first part of the analysis identified amino acid L at position 11 as differing in frequency between two groups, therefore we explored its association with two phenotypes. **There were 56 patients with L at position 11**. Thirty-seven of them (67,2%) had not-fibrotic phenotype (group 2), while only 18 of the remaining 51 subjects which carried G at position 11 had not-fibrotic phenotype (group 2); this difference was statistically significant ($p=0.0019$). The 56 patients positive for amino acid L at position 11 overlapped to a great extent with the 56 who had the DEDEAV combination at positions 55,56, 84-87, but not completely. **There were 49 subjects with L at position 11 who had concurrently DEDEAV motif**, while 7 had other combinations.

b) HLA – DQB1

Regarding the HLA - DQB1 analysis, there were 104 patients with data available. There were 632 possible combinations of two amino acids and 629 of them were present in the study population. **Amino acid G at position 26**, produced a difference in the first part of the analyses. It was present in 51 patients of whom 35 (69%) had non-fibrotic

phenotype (group 2), compared to 20 out of 53 patients (38%) with other amino acids at position 26 who presented non-fibrotic phenotype ($p=0.0018$). The same 51 patients had amino acid **S at position 74**, with the same distribution between the study groups (69% vs 38%, $p = 0.0018$).

c) HLA – DRB1

One hundred five patients had available data for amino acidic sequence of HLA - DRB1. There were 923 possible combinations of pairs of amino acids, and 829 of them were present in the study population. Amino acid **F at position 67** was present in 62 patients; 41 of them (66%) had non-fibrotic phenotype (group 1) versus 14 out 43 (33%) of patients with other amino acids at the same position ($p=0.0008$). **Combination of F at 67 with amino acid D at position 70** reallocated one patient out of the 62, making this difference 41/61 (67%) versus 14/44 (32%), with $p=0.0004$.

4.3. Association of triplets of amino-acids with risk of developing fibrotic phenotype

a) HLA – DPB1

There were 5164 possible amino acid triplets among the 14 **DPB1 positions**; 4530 of them were present in our study population. When they were tested for the association with the fibrotic/non fibrotic phenotype, the strongest association was observed between **two triplets having L, A and A at positions 11, 36 and either 55 or 56** and non-fibrotic phenotype. LAA (11,36,55) and LAA (11,36,56) were present in the same 40 patients. Twenty-nine out of these 40 patients (72%) had non-fibrotic phenotype (group 2) versus 26 (39%) of the 67 patients who did not present the triplets ($p=0.0012$). **These triplets were very strongly correlated with the LVDEDEAV pattern at positions 11, 36, 55-56 and 84-87**. Eighty-four out of 107 patients had the concurrent presence of both triplets and LVDEDEAV combination.

The second strongest DPB1 triplet association with the non-fibrotic phenotype (group 2) was shared by a set of **21 triplets** that were all sub-sequences of a combination of 10 specific amino acids across ten positions (VFVDEIDEAV at positions 8,35,36,55,56,65,84-87) ($p=0.0019$). Notably, **seven of these 10 amino acids were represented by the VDEDEAV** portion of the previous LVDEADEAV DPB1 combination at positions **36, 55-56 and 84-87**. The **other three were amino acids V, F and I at positions 8, 35 and 65**. There were 54 patients with at least one of these triplets; 49 of them had simultaneously LVDEDEAV combination and remaining 5 had other motifs.

Since Inamoto et coll. reported that the risk of sclGVHD was decreased in patients with GPM motifs at positions 85-87 of the DPB1 (**279**), we also explored the distribution of this motif in our study population. This combination was present in 91 of 107 patients, but it was perfectly balanced between two groups. It was present in 44 out of 52 patients (84,6%) in group 1 versus 47 out of 55 patients (85,5%) in group 2 ($p=1.0$).

b) HLA – DQB1

There were 6606 possible triplets within the HLA – DQB1; 6069 of them were present in our study population. When they were tested for the association with the fibrotic/non fibrotic phenotype, the strongest association was shared by **36 triplets** which were present in 35 out of the 55 patients (64%) with non-fibrotic phenotype versus 16 out of the 49 patients (33%) with fibrotic phenotype (group 1), ($p=0.0018$). From the other perspective, 35 out of the 51 patients (69%) with the triplets had non-fibrotic phenotype, while only 20 (38%) of the 53 patients without the triplets had not-fibrotic phenotype.

Further examination of the triplets revealed that **they were all sub-sequences of a combination across eight positions, 26, 28, 37, 52, 55, 56, 74 and 75, containing amino acids G, T, Y, P, R, P, S and V**. Indeed, all 51 patients with any of the 36 triplets had this entire combination present. This may limit the specificity of the association because it is possible that the entire combination of eight amino acids is implicated rather than one of the triplets or some other combinations of four, five, six or seven amino acids.

In order to confirm the association of this combination with the non-fibrotic phenotype (group 2) we changed the outcome from the simple present/absent dichotomy to the number of the eight amino acids present in each patient, which ranged from 2 to 8 (Table 4.6). Two statistical tests, based on the ranks of the ordered values of the difference between the two study groups in these numbers, confirmed the association, raising the statistical significance to $p=0.0010$ and $p=0.0007$.

Table 4.6 Comparison of numbers of amino acids matching the DQB1 GTYPRPS combination between the study groups.

		Number of amino acids matching DQB1 combination					
		2	4	5	6	8	Total
Group 1	N	3	9	16	5	16	49
	%	6	18	33	10	33	100
Group 2	N	0	3	16	1	35	55
	%	0	5	29	2	64	100
Total	N	3	12	32	6	51	104

N = number of patients.

c) HLA – DRB1

There were 11429 possible triplets from the amino acids observed at 14 **DRB1 positions**, 8020 of them were present in our study population. The most significant association with the study groups consisted in the association of non-fibrotic phenotype (group 2) with **five triplets which had amino acids F and D at positions 67 and 70** ($p<0.00042$).

Slightly less significant was the association between the triplets formed by amino acid **R at position 71, A at position 73 and either Y at position 10 or S at position 11** (YRA at 10,71,73 and SRA at 11,71,73) and non-fibrotic phenotype. These triplets were

simultaneously present or absent in any of the patients and their occurrence was associated with non-fibrotic phenotype (p=0.0007). **The patients with this combination overlap to a great extent with the patients with F, D at positions 67 and 70;** in fact, 84 out of 107 patients had the simultaneous presence or absence of these two amino acidic combinations.

HLA- DPB1, DQB1 and DRB1 association with fibrotic/non fibrotic phenotypes are shown in Table 4.7, Table 4.8 and Table 4.9.

Table 4.7 HLA- DPB1 associations

V	L	F	V	D	E	I	D	E	A	V	L vs (D1 + sclGVHD)	p
8	11	35	36	55	56	65	84	85	86	87		
-	L	-	-	-	-	-	-	-	-	-	67% vs 37%	0.0019
V	-	-	V	D	-	-	-	-	-	-	65% vs 35%	0.0019
V	-	-	V	-	E	-	-	-	-	-	65% vs 35%	0.0019
-	L	-	A	A	-	-	-	-	-	-	53% vs 21%	0.0012
-	L	-	A	-	A	-	-	-	-	-	53% vs 21%	0.0012
-	-	-	-	D	E	-	D	E	A	V	67% vs 37%	0.0019
L	-	V	D	E	-	D	E	A	V		67% vs 37%	0.0019
21 triplets derived from VFVDEIDEAV (8,35,36,55,56,65,84 - 87)											67% vs 37%	0.0019

Table 4.8 HLA- DQB1 associations

G 26	T 28	Y 37	P 52	R 55	P 56	S 74	V 75	L vs (D1 + sclGVHD)	p
G	-	-	-	-	-	-	-	64% vs 33%	0.0018
-	-	-	-	-	-	S	-	64% vs 33%	0.0018
G	-	-	-	-	-	S	-	64% vs 33%	0.0018
36 triplets derived from GTYP RPSV (26, 28, 37, 52, 55, 56, 74,75)								64% vs 33%	0.0018

Table 4.9 HLA- DRB1 associations

10 Y	11 S	67 F	70 D	71 R	73 A	L vs (D1 + sclGVHD)	P
-	-	F	-	-	-	75% vs 42%	0.0008
-	-	F	D	-	-	75% vs 40%	0.0004
-	S	-	-	R	A	70% vs 40%	0.0007
Y	-	-	-	R	A	70% vs 40%	0.0007
5 triplets with F and D at positions 67 and 70						75% vs 40%	0.0004

Subsequently, we tested the joint associations of amino acidic combinations derived from different HLA genes with fibrotic/non fibrotic phenotype. We included in the analysis those sequences that have been found to be correlated singularly with fibrotic/non fibrotic phenotype in our study population.

Since HLA class II alleles are inherited with strong linkage disequilibrium, we tested if there was an association between these amino acidic sequences, deriving from different HLA genes. No association was found between LVDEDEAV (11,36,55,56,84-87) in DP1 and LGTYPRPSV (26, 28, 37, 52, 55, 56, 74,75) in DQB1, and among these combinations and DRB1 triplets.

We performed a logistic regression analysis with amino acidic sequences from each HLA gene.

Regarding DPB1, we considered LAA triplets (11,35, 55/56), LVDEDEAV (11,36,55,56,84-87), VVDEIDEAV (8,35,36,55,56,65,84 - 87), and VDEDEAV (36, 55, 56, 84-87), which presented the overlap of the previous three combinations. We considered GTYPRPSV (26, 28, 37, 52, 55, 56, 74,75) sequence for DQB1, and FD at positions 67 and 70 for DRB1. In this way we generated four logistic models with three factors each, one factor from each gene. The exact p values of the joint significance of the models range from $p = 6.2e-6$ to $p = 1.6e-5$ (Table 4.10). The combination with the smallest p value was the one with VDEDEAV (36, 55, 56, 84-87) in DPB1.

Table 4.10 Combined associations of amino acidic sequences derived from three HLA genes and non-fibrotic phenotype

DPB1	LVDEDEAV 11,36,55,56,84,85,86,87	LAA 11,36,55/56	VVDEIDEAV 8,35,36,55,56,65,84-87	VDEDEAV 36,55,56,84-87
DQB1	GTYPRPSV 26,28,37,52,55,56,74,75	GTYPRPSV 26,28,37,52,55,56,74,75	GTYPRPSV 26,28,37,52,55,56,74,75	GTYPRPSV 26,28,37,52,55,56,74,75
DRB1	FD 67,70	FD 67,70	FD 67,70	FD 67,70
p	1.6e-5	1.6e-5	1.6e-5	6.2e-6

We also inserted DRB1 triplets SRA (10,71,73) and YRA (11,71,73) as the DRB1 factor in the logistic model with three factors, but these tests showed no improvement over the best models previously found.

4.4. Qualitative and quantitative analysis of amino acids associated with fibrotic/non fibrotic phenotype

In the third part of the analysis we performed qualitative and quantitative analysis of amino-acids associated with fibrotic/non fibrotic phenotype.

Differential analysis of the amino acids grouped for:

- a) Family (polar/non polar with 4 subfamilies)
- b) Similarity of the strength of the H bonds that they can harbour in their vicinity according to Rogov et al (285) was performed between the group 1 (fibrotic phenotype) and group 2 (non-fibrotic phenotype).

4.4.1. Qualitative analysis of amino-acids

The amino acids were divided in classical categories (families), according to their physical, chemical and structural properties, defined as:

1= Polar only (contains: W, Y, Q)

1A = Polar + small/very small (contains: N, S, C, T)

1AB = Polar + charged + small (contains D)

1B = Polar + charged only (contains E)

1B1 = polar + charged + positive (contains: R, K, H)

2 = Non polar, only (contains: F, M, I, L)

2A = non polar + small/very small (contains: V, A, G, P)

Some of the positions had multiple amino acids belonging to the same family, thus the number of positions tested was reduced to 11 in HLA - DPB1, 12 in HLA - DQB1 and 11

in HLA - DRB1. After Hochberg correction, **DPB1 position 11** was the only one with a **significant difference** between two study groups **in the family frequencies** ($p=0.033$). When the **FDR correction** was applied, differences in the family frequencies for **the HLA - DQB1 position 26** resulted also significant.

Since the most important features for epitope binding are polarity and hydrophobicity, in order to reduce the number of categories tested with consequent improvement of the power of the analysis, we reclassified amino acids as follows:

- Hydrophobic (M, I, V, L, H, Y, W, F, K, C, T, A) and Not-hydrophobic (G, S, D, E, R, P, N, Q)
- Polar (Q, N, S, C, T, Y, W, H, K, R, D, E) and Not-polar (P, G, A, I, V, L, M, F). We subsequently divided polar amino acids in positive charged (R, K, H), negative charged (D, E) and not-charged (Q, N, S, C, T, Y, W).

When the amino acids were grouped as **hydrophobic versus non-hydrophobic**, 36 of the 61 positions displayed only one type of amino acids (20 hydrophobic and 16 non-hydrophobic), leaving 25 with a difference to test. The only difference between the study groups that remained significant even after the less stringent FDR correction was at **DRB1 position 71** ($p < 0.05$). The frequency of hydrophobic amino acids in group 1 (fibrotic phenotype) was of 34%, compared to 12% in the group 2 (non-fibrotic phenotype).

When the amino acids were grouped as **polar versus non-polar**, 41 of the 61 positions had only one type of amino-acids represented (22 polar and 19 non-polar), and 20 had a difference to test. Four of the 20 had individual p values slightly less 0.05, but none of them was significance after correction. Nevertheless, the largest difference was observed for the position 56 of DPB1, although not statistically significant.

We also tested the combined frequencies of **hydrophobic and non-polar** (A, F, I, L, M, V) amino acids **versus the combined frequencies of all the others**. The only two positions that had significant differences after the Hochberg correction were:

- **DPB1 position 11** and (there were 41% of patients with hydrophobic and non-polar aminoacids in group 2 versus 20% of patients in group 1, $p=0.026$)

- **DQB1 position 26** (there were 35% of patients with hydrophobic and non-polar aminoacids in group 1 versus 57% in group 2, $p=0.047$).

The classifications of hydrophobic versus non-hydrophobic and polar versus non-polar did not produce differences that reached the above level of statistical significance.

4.4.2. Quantitative analysis of amino-acids

For the quantitative analysis using the coefficients of similarity suggested by Rogov (285), we identified the sequences expressed differently between two groups of patients in a significant manner and selected the longest one (one for every of the three HLA genes) as a reference sequence.

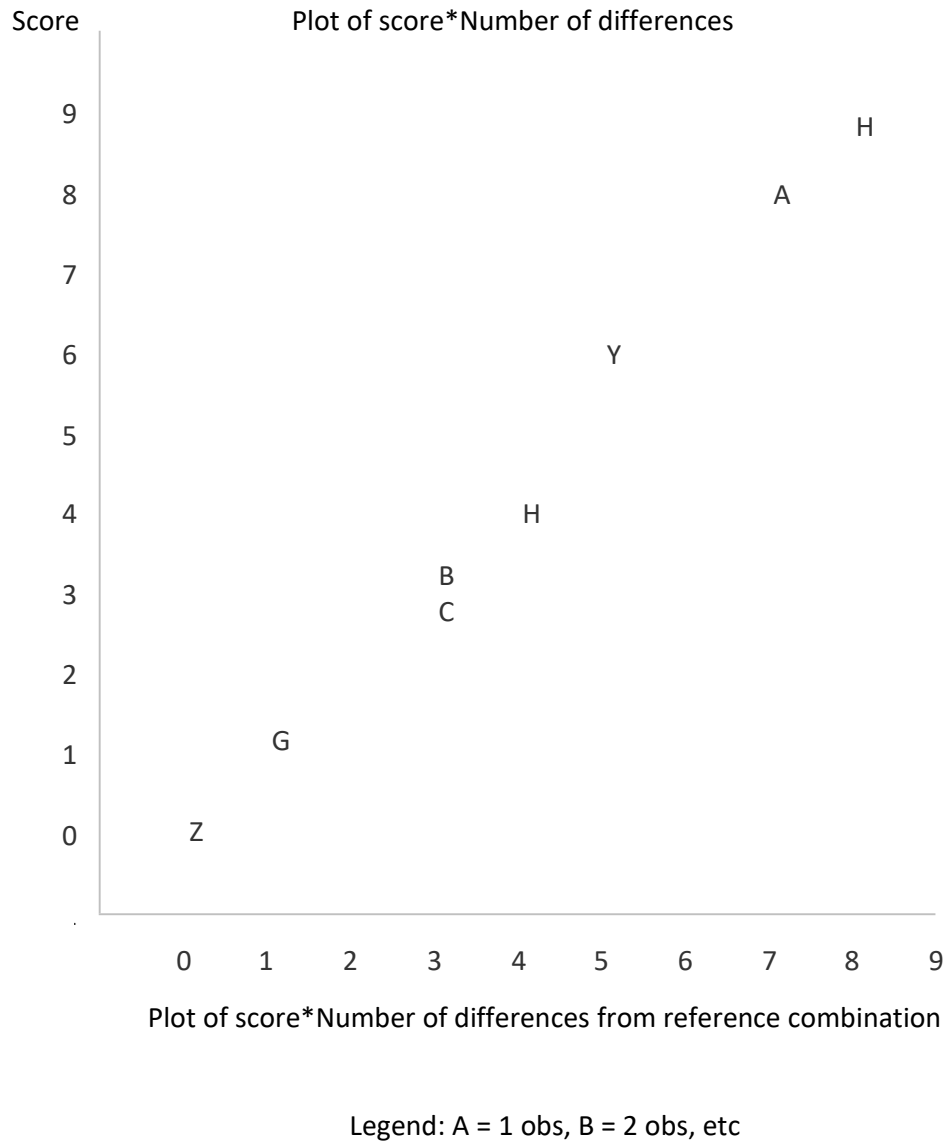
Regarding DPB1, the sequence L-V-D-E-D-A-V-T referring to positions 11,36,55,56,84-87 was taken as reference for the respective positions. Graph. 4.1 shows the diversity scores. As indicated in the legend at the top of the page, the plotting symbols are the letters of the alphabet corresponding to the numbers of patients with equal scores. The H in the upper right corner of the graph represents the 8 patients with scores of 8,804 and the A represents one patient with score of 8,036. The Z with coordinates (0,0) represents 49 patients with the score of 0 (these patients had at least one reference sequence). The G represents 7 patients with scores of 1,205. The C and B represent 7 patients with scores of 2,772 and 2 patients with scores of 3,129, respectively. The H and Y represent 8 patients with scores of 3,977 and 25 patients with scores of 6,032 respectively.

The group 2 (non-fibrotic phenotype) had a significant tendency to have fewer differences from the reference sequence than the group 1 (fibrotic phenotype) ($p=0.0007$).

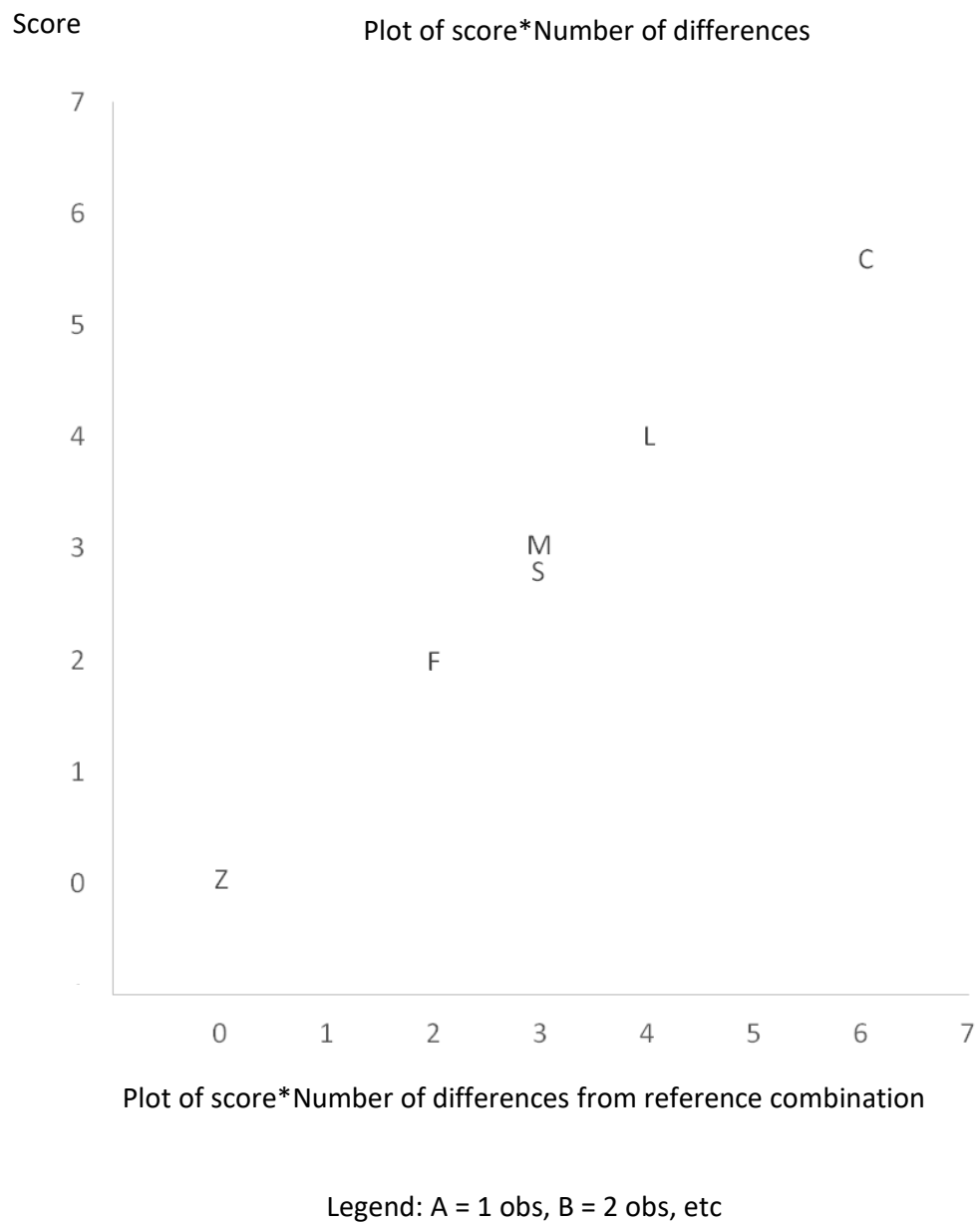
For the DQB1 we used the reference sequence the amino acids G, T, Y, P, R, P, S and V at DQB1 positions 26, 28, 37, 52, 55, 56, 74 and 75. Graph. 4.2 shows the diversity scores. The S in the centre of the plot represents 13 patients with scores of 2,877 and 6 with scores of 2,904, the M above it represents 12 patients with scores of 2,981 and 1 with

3,008, and the L to the right and higher up represents 10 with 3,929 and 2 with 3,956. The F and the C represent patients with scores of 1,956 and 5,513, respectively. The Z presents 51 patients with at least one reference sequence (score 0). The tendency of group 2 (non-fibrotic phenotype) to have fewer differences from the reference sequence than the group 1 (fibrotic phenotype) was statistically significant also for DQB1. ($p=0.0010$).

Graph. 4.1 Diversity scores and numbers of differences from the reference combination (DPB1)



Graph. 4.2 Diversity scores and numbers of differences from the reference combination (DQB1)



We did not obtain significant results regarding DRB1.

5. DISCUSSION

This is the first study which describes common sequences (shared epitopes), codified by three of the most important genes associated with SSc: HLA – DPB1, HLA – DQB1 and HLA – DRB1, that seem to display a protective effect against severe fibrotic involvement in SSc and sclGVHD. We identified common amino acidic sequences within the hypervariable regions of β chains of HLA – DP, HLA – DQ and HLA – DR molecules present in lcSSc patients with minimal skin fibrosis but not in SSc and sclGVHD patients with severe skin fibrosis.

A few studies addressed the shared epitope hypothesis in SSc but were focused on the role of common sequences in the susceptibility towards SSc and derived from the comparison of SSc patients and healthy controls. No study has investigated the common amino-acidic sequences between SSc and sclGVHD so far.

Azouz et al analyzed the HLA-DQB1 and HLA-DRB1 susceptibility alleles for SSc in a cohort of French Caucasian patients (267). They identified two common amino acidic motifs codified by these alleles: FLEDR at positions 67 – 71 of the HLA-DR β chain and TRAE LDT at positions 71 – 77 of the HLA-DQ β chain (267). When the patients were divided by clinical subset, both sequences were associated only with diffuse SSc. When the autoantibody status was considered, the association was confirmed only in ATA positive dSSc patients. The addition of a tyrosine at the position 30 strengthened the TRAE LDT association with dSSc (267). Moreover, the risk of ATA positive dSSc doubled in case of double presence of the shared epitope (267).

We did not observe different distribution of these amino acidic motifs reported in our study population. It is important to note that our study population was different from the population studied by the French authors. First of all, in the study performed by Azzouz et al, the association of the shared epitopes with dSSc was only present in ATA positive patients, suggesting that FLEDR and TRAE LDT motifs might be associated with ATA production, rather than with clinical subset and with the extension of skin fibrosis. In the present study we analysed the clinical outcome (severe skin fibrosis) in patients

with two different fibrotic diseases. Only in SSc the extent of skin fibrosis is correlated with ATA production whilst autoantibodies have not been demonstrated to play a role in the pathogenesis of sclGVHD (**171**). It is clear that potential common pathogenic mechanism underlying fibrosis in SSc and sclGVHD is not related to ATA production, and this may completely explain the apparent discrepancy of our results with those reported by Azzouz. To note that in our study population 80% of dSSc patients were ATA positive and none of sclGVHD patient were ATA positive.

In addition, Azzouz et al classified patients according to the presence of skin involvement proximal or distal to elbows and knees, whilst we selected patients according to the degree of skin involvement (quantified by mRSS in SSc and percentage of BSA covered by fibrotic skin in sclGVHD). Mean values of mRSS have not been reported in the paper published by Azzouz, therefore it is not possible to compare the severity of skin fibrosis between our study population and those described by the French authors.

The association of HLA-DQB1 alleles that encode a non-leucine residue at position 26 (HLA-DQB1 26 epi) with ACA positive SSc has been reported in US and UK Caucasian and Chinese Han patients (**250, 260, 272**). On the contrary, in African Americans HLA-DQB1 26 epi was associated with ATA positive SSc (**250**). In our study glycine at position 26 was associated with ISSc. We have not performed the subanalysis in the limited SSc ACA positive patients, but in our study population more than 80% of ISSc patients were ACA positive.

Recently Mayes et al have performed a large study on 5850 SSc patients and 9401 healthy controls of European origin, genotyped with the ImmunoChip array (**254**). Classical alleles, amino acid residues, and SNPs across the HLA region were imputed and tested (**254**). The authors obtained a model composed of six polymorphic amino acid positions and seven SNPs which successfully conditioned all observed associations in the HLA region with SSc and its serological subsets (**254**). Amino acids F and V at positions 67 and 86 of HLA – DR β chain, I at position 76 of HLA – DB β chain and absence of amino acids K and R at 96 of HLA – DB β chain were risk factors for ATA positive SSc, while the protective amino acid was represented by L at position 67 of HLA – DR β chain (**254**). Regarding ACA positive SSc, F at position 13 of HLA – DR β chain and T at position 69 of

HLA – DQ α were associated with risk for the disease. Amino acids Y at position 13 of HLA – DR β and L at 69 of HLA – DQ α chain conferred protection (**254**).

Differently from Mayes et al, we observed the protective effect of amino acid F at position 67 against dSSc and sclGVHD. As previously mentioned, our study population comprised also patients with sclGVHD in which ATA do not play a pathogenic role. Amino acids at position 13 were not distributed differently between the two groups in our study. To note that Mayes et al compared SSc patients to healthy subjects and the methodology for HLA typing was different between the two studies (Immunochip versus high resolution HLA typing). Amino acids 76 and 96 of β chains of HLA – DP and amino acid 86 of HLA – DR molecule are not the part of their binding pockets (**36, 287**), thus they have not been investigated in our study, as well as amino acids codified by HLA-DQA1.

There is only one report (congress presentation) that investigated the role of HLA in genetic susceptibility towards sclGVHD (**279**). Inamoto et al investigated the role of 15 SNPs previously reported to be associated with SSc susceptibility and located within the following genes: HLA- DQB1, HLA – DP, HLA-DRA, CD247, BLK (B lymphocyte kinase), BANK1 (B-cell scaffold protein with ankyrin repeats 1), IL12RB2 (Interleukin 12 receptor subunit beta 2), IRF8 (IFN regulatory factor 8), PTPN22 (Protein tyrosine phosphatase, non-receptor type 22), STAT4, TNIP1 (TNFAIP3-interacting protein 1), TNPO3-IRF5 (Transportin 3 - IFN regulatory factor 5) and TNFSF4 (Tumor necrosis factor ligand superfamily member 4) (**279**). The analysis revealed potential role of HLA-DP SNPs, and further exploration of the association between HLA-DP and sclGVHD showed the protective role of amino acidic GMP motif at positions 85-87 of the β chain of HLA- DP molecule and of amino acid Q at position 31 of the HLA- DP α chain (**279**). We explored the distribution of GMP motif (85-87) motif in our study population. This combination was equally distributed in ISSc patients and in patients with dSSc and sclGVHD, who presented severe skin fibrosis.

The key difference of our study with aforementioned studies on the shared epitope in SSc is that our aim was not to find an amino acidic sequence that predispose for SSc or

for sclGVHD, but for the severe fibrotic involvement in course of both diseases. For this reason we did not perform comparative analysis between patients and healthy subjects. Importantly, the shared epitopes seem to play a role in ATA production in SSc, as previously mentioned.

In SSc, the presence of specific autoantibodies is strongly associated with distinct clinical phenotypes, but there is no unanimous consensus on their direct role in the pathogenesis of SSc or its clinical manifestations (**288**). Nevertheless, Fava et al have recently shown that the frequency of circulation DNA topoisomerase I specific CD4 T cells predicts presence and progression of interstitial lung disease in SSc (**289**). In addition, it has been reported that T cell lines from SSc patients and healthy controls recognize different epitopes on DNA topoisomerase I (**290**). It might be hypothesized that shared epitopes on HLA molecules bind specific portions of DNA topoisomerase and present them to T cells. To note that not all SSc patients and not all patients with diffuse SSc patient are ATA positive, and that other mechanism might be involved in pathogenesis of the disease and of its fibrotic manifestations.

On the other hand, no association between autoantibodies and specific phenotypes of cGVHD has been reported so far and there is no role for ATA in the pathogenesis of sclGVHD and its fibrotic manifestations (**171**).

The most interesting finding of our study is the existence of common amino acidic motifs protective against development of severe skin fibrosis shared by two similar, thus different diseases, both characterized by immune-mediated fibrosis. Moreover, the simultaneous presence of the shared motifs at different HLA molecules (DP, DQ and DR) conferred the higher degree of protection against skin fibrosis. These sequences are located in hypervariable regions of β chains of HLA-DP, DQ and DR molecules.

We hypothesized common pathogenic pathway of immune-mediated fibrosis in sclGVHD and dSSc. Given the role of these regions in binding the antigenic peptides and presenting them to T lymphocytes for antigen recognition, we may hypothesize that the HLA molecules which share common amino acidic motifs, found in our study, might be capable of binding the same/functionally similar peptides which may produce anti-

fibrotic effects or cause down-regulation of fibrosis during further autoimmune response. Another possibility is that the HLA molecules containing the shared epitopes might not be able to bind specific auto-antigenic peptides that trigger a pro-fibrotic autoimmune response, thus preferentially expressing other pathogenic pathways of both diseases. The fact that the simultaneous presence of the shared motifs at different HLA molecules seem to confer the higher degree of protection against skin fibrosis, may suggest an interplay of different autoantigens in determining fibrotic response.

Once the peptidic antigens bind to peptide groove of HLA molecules they are presented to T cells for antigen recognition. T cells play a critical role in mediating the response to antigens in both SSc (**123**) and sclGVHD (**178**). In SSc, differently from chronic GVHD, there are no identified auto-antigens, but it is accepted that in both diseases there is a breakdown in self-tolerance leading to T cells activation and secretion of pro-inflammatory and fibrogenic cytokines (**123, 178**). In addition, no specific antigens have been implicated in the pathogenesis of sclGVHD, compared to other forms of cGVHD. An appealing hypothesis is that there is a common autoantigen/s which interacts with specific T cells that exert protective effect against immune mediated fibrosis.

We also performed qualitative and quantitative analysis of amino acids contained in the shared epitope and hypothesized that functionally similar motifs may play the same role in the pathogenesis of immune-mediated fibrosis. The qualitative analysis of amino acids, classified according to their physical, chemical and structural properties, underlined the importance of hydrophobicity in association with different fibrotic phenotypes. The presence of a hydrophobic amino acidic residue at DRB1 position 71 was significantly associated with development of severe skin fibrosis in SSc and sclGVHD. The role of hydrophobicity as the major determinant of the number of peptide bonds that each amino acid is able to form with other amino acids in its surrounding has been already underlined by Rogov whose method for quantitative analysis was applied in our study (**285**). To the best of our knowledge this is the first study that applied the Rogov's coefficients of similarity between the amino acids, based on the number of peptide bonds that each amino acid is able to form with other amino acids in its surrounding, for the estimation of the binding properties of the given peptide. The quantitative analysis,

based on a degree of similarity of the peptidic sequences, showed that patients who did not develop fibrotic phenotype had either shared epitope or peptidic sequences functionally similar to the shared epitope, whilst patients with fibrotic phenotype had peptides functionally different from the shared epitope.

The shared motifs may be used as predictive factors of severity in scl-GVHD or SSc. It is known that the extent of skin fibrosis is an indicator of severity of SSc and correlates with internal organ involvement and survival in SSc (**99, 280**).

Skin fibrosis cause significant morbidity and disability also in sclGVHD (**174, 176**). In our study, patients with severe skin fibrosis carried less frequently the shared epitopes compared to patients with minimal fibrosis. Therefore, the patients who present one or more shared sequences might be at lower risk of developing severe disease, high morbidity and poor outcome. The presence of these sequences could be tested to stratify patients for risk of having/not having severe disease who need/do not need close monitoring and aggressive treatment in future clinical trials.

Anti-topoisomerase antibodies are already used as independent predictor of more severe SSc with diffuse cutaneous fibrosis (**291**), however not all ATA positive patients develop severe skin fibrosis and severe SSc and do not need potentially toxic therapies. In SSc, shared epitope might be used in addition to the absence of ATA as a benign prognostic factor. On the other hand, given the absence of any other prognostic factor in sclGVHD patients, the potential use of the shared epitope as a prognostic factor for sclGVHD severity is intriguing.

Another interesting finding of our study is that 20 out of 61 investigated positions across the hypervariable regions of β chains of three HLA molecules had the same amino acids position at both chromosomes, in all patients. It will be very interesting to compare amino acids at these positions with those in healthy subjects, in order to assesses their potential role in the pathogenesis and risk of developing both diseases.

Nevertheless, our study has some important limitations. First of all, there were differences in the genetic background of SSc and sclGVHD patients. All SSc patients had Caucasian European origin and had Italian ethnicity. On the contrary 80% of sclGVHD patients had Caucasian American origin, whilst 20% had other ethnical background

(African American, Hispanic and Asiatic). Since HLA associations with the susceptibility to SSc seem to be different in different races and populations (**250**), different genetic background of two study populations might have influenced the results. However, a large study on outcomes after HSCT in different US ethnical groups, did not observe any significant difference in the risk of developing cGVHD between whites and other ethnic groups (African Americans, Hispanic and Asiatic) (**292**).

Another important limitations are: the small number of subjects studied and the fact that we did not analyse extensively possible linkages between the amino acids derived from three HLA genes. It is well known that HLA class II alleles are inherited with strong linkage disequilibrium, so it is possible that the association of one or more shared motifs with clinical outcome is related to the linkage with other alleles which codifies for other shared epitopes associated with clinical phenotypes.

It has been suggested in RA, that HLA associations with the disease may differ depending on the stage of the disease (**58**). We have not explored this possibility in our study.

However, the biggest limitation consists in the fact that we presumed the common pathogenic pathway of immune-mediated fibrosis in SSc and sclGVHD. Although sclGVHD reassembles some of the clinical features of SSc, the most important difference consists in the fact that sclGVHD is directly related to the transplant procedure and is mediated by donor's T cells reactive against recipient's antigens (**178**), whilst the exact etiological agent of SSc is still unknown. In SSc, the autoimmune response plays a fundamental role in its pathogenesis and the specific autoantibodies characterize the disease (**97**). In sclGVHD, there is a loss of self tolerance and low titer autoantibody production, but it is not an autoimmune condition (**178**).

It is also possible that genetic and epigenetic mechanisms, other than HLAs, regulate the expression and severity of fibrosis in SSc and sclGVHD.

Our results have to be validated on independent large cohorts of patients with equivalent characteristics and outcomes. Thus, further studies on large populations are necessary to confirm our preliminary findings.

6. CONCLUSIONS

This is the first study which describes common sequences (shared epitopes), codified by three of the most important genes associated with SSc: HLA – DPB1, HLA – DQB1 and HLA – DRB1, that seem to display a protective effect against severe fibrotic involvement in SSc and sclGVHD. We identified common amino acidic sequences within the hypervariable regions of β chains of HLA – DP, HLA – DQ and HLA – DR molecules present in patients with minimal skin fibrosis but not in patients who developed severe skin fibrosis. Additionally, we proposed to quantify functional differences within the HLA motifs to compare apparently different sequences for their functional ability of binding specific peptide residues. Our findings indicate this approach can add value to simple genotyping of HLA. Further studies on large populations are necessary to confirm our preliminary findings.

7. REFERENCES

1. **Mehra NK**, Kaur G, Histocompatibility Antigen Complex of Man. 2016 In: eLS. John Wiley & Sons Ltd, Chichester. <http://www.els.net> [doi: 10.1002/9780470015902.a0001234.pub4]
2. **Trowsdale J**, Knight JC. Major histocompatibility complex genomics and human disease. *Annu Rev Genomics Hum Genet.* 2013;14:301-23.
3. **Hung SI**, Chung WH, Jee SH, Chen WC, Chang YT, Lee WR, Hu SL, Wu MT, Chen GS, Wong TW, Hsiao PF, Chen WH, Shih HY, Fang WH, Wei CY, Lou YH, Huang YL, Lin JJ, Chen YT. Genetic susceptibility to carbamazepine-induced cutaneous adverse drug reactions. *Pharmacogenet Genomics.* 2006;16(4):297-306.
4. **Dainichi T**, Uchi H, Moroi Y, Furue M. Stevens-Johnson syndrome, drug-induced hypersensitivity syndrome and toxic epidermal necrolysis caused by allopurinol in patients with a common HLA allele: what causes the diversity? *Dermatology.* 2007;215(1):86-8.
5. **Shiina T**, Hosomichi K, Inoko H, Kulski JK. The HLA genomic loci map: expression, interaction, diversity and disease. *J Hum Genet.* 2009;54(1):15-39.
6. **Garnier JM**, Sharp PP, Burns CJ. BET bromodomain inhibitors: a patent review. *Expert Opin Ther Pat.* 2014;24(2):185-99.
7. **Abbas A**, Lichtman A, Pillai S. (2014) Major histocompatibility complex molecules and antigen presentation to T lymphocytes in Abbas A, Lichtman A, Pillai S (Eds.), *Cellular and Molecular Immunology* (pp 107-136). Philadelphia: WB Saunders.
8. **Murphy K**. Antigen Presentation to T Lymphocytes (2011) in Murphy K (Eds.) *Jeneway's Immunobiology*. New York: Garland Science.
9. **Blum JS**, Wearsch PA, Cresswell P. Pathways of antigen processing. *Annu Rev Immunol.* 2013;31:443-73.
10. **Klein J**, Sato A. The HLA system. First of two parts. *N Engl J Med.* 2000;7;343(10):702-9.

11. **Roche PA**, Teletski CL, Stang E, Bakke O, Long EO. Cell surface HLA-DR-invariant chain complexes are targeted to endosomes by rapid internalization. *Proc. Natl. Acad. Sci. USA* 1993;90:8581–85.
12. **Ghosh P**, Amaya M, Mellins E, Wiley DC. The structure of an intermediate in class II MHC maturation: CLIP bound to HLA-DR3. *Nature* 1995;378:457–62.
13. **Yoon T**, Macmillan H, Mortimer SE, Jiang W, Rinderknecht CH, Stern LJ, Mellins ED. Mapping the HLA-DO/HLA-DM complex by FRET and mutagenesis. *Proc Natl Acad Sci U S A*. 2012;109(28):11276-81.
14. **Madden DR**. The three-dimensional structure of peptide-MHC complexes. *Annu Rev Immunol*. 1995;13:587-622.
15. **Warren EH**, Zhang XC, Li S, Fan W, Storer BE, Chien JW, Boeckh MJ, Zhao LP, Martin PJ, Hansen JA. Effect of MHC and non-MHC donor/recipient genetic disparity on the outcome of allogeneic HCT. *Blood*. 2012 Oct 4;120(14):2796-806.
16. **Schueler-Furman O**, Altuvia Y, Margalit H. Examination of possible structural constraints of MHC-binding peptides by assessment of their native structure within their source proteins. *Proteins* 2001; 45: 47–54.
17. **Stern LJ**, Brown JH, Jardetzky TS, Gorga JC, Urban RG, Strominger JL, Wiley DC. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature*.1994;368(6468):215–21.
18. **Stern LJ**, Wiley DC. Antigenic peptide binding by class I and class II histocompatibility proteins. *Structure*. 1994;15;2(4):245-51.
19. **Jones EY**, Fugger L, Strominger JL, Siebold C. MHC class II proteins and disease: a structural perspective. *Nat Rev Immunol*. 2006;6(4):271-82.
20. **James EA**, Moustakas AK, Bui J, Nouv R, Papadopoulos GK, Kwok WW. The Binding of Antigenic Peptides to HLA-DR Is Influenced by Interactions between Pocket 6 and Pocket 91; *J Immunol*. 2009; 183(5): 3249–3258.
21. **Ferrante A**. For Many But Not For All: How The Conformational Flexibility Of The Peptide/Mhcii Complex Shapes Epitope Selection. *Immunol Res*. 2013; 56(1): 85–95.

22. **Jameson L. J.** Principles of Molecular Medicine (1998). Springer Science & Business Media.
23. **Zoete V**, Irving M, Ferber M, Cuendet MA, Michielin O. Structure-Based, Rational Design of T Cell Receptors. *Front Immunol.* 2013;12;4:268.
24. **Reinherz EL**, Wang JH. Codification of bidentate pMHC interaction with TCR and its co-receptor. *Trends Immunol.* 2015;36(5):300-6.
25. **Rangarajan S**, Mariuzza RA. T cell receptor bias for MHC: co-evolution or co-receptors? *Cell Mol Life Sci.* 2014;71(16):3059-68.
26. **Lee JK**, Stewart-Jones G, Dong T, Harlos K, Di Gleria K, Dorrell L, Douek DC, van der Merwe PA, Jones EY, McMichael AJ. T cell cross-reactivity and conformational changes during TCR engagement. *J Exp Med.* 2004, 6;200(11):1455-66.
27. **Kass I**, Buckle AM, Borg NA Understanding the structural dynamics of TCR-pMHC complex interactions. *Kass Trends Immunol.* 2014;11;35(12):604-612.
28. **Borbulevych OY**, Piepenbrink KH, Baker BM. Conformational melding permits a conserved binding geometry in TCR recognition of foreign and self molecular mimics. *J Immunol.* 2011;1;186(5):2950-8.
29. **Sanchez-Mazas A**, Vidan-Jeras B, Nunes JM, Fischer G, Little AM, Bekmane U, Buhler S, Buus S, Claas FH, Dormoy A, Dubois V, Eglite E, Eliaou JF, Gonzalez-Galarza F, Grubic Z, Ivanova M, Lie B, Ligeiro D, Lokki ML, da Silva BM, Martorell J, Mendonça D, Middleton D, Voniatos DP, Papasteriades C, Poli F, Riccio ME, Vlachou MS, Sulcebe G, Tonks S, Nevessignsky MT, Vangenot C, van Walraven AM, Tiercy JM. Strategies to work with HLA data in human populations for histocompatibility, clinical transplantation, epidemiology and population genetics: HLA-NET methodological recommendations. *Int J Immunogenet.* 2012;39(6):459-72.
30. **Howel WM**, Poole KL (2002) The Special case of HLA genes: detection and resolution of multiple polymorphic sites in a single gene in Day INM. (Ed.) Molecular genetic epidemiology. A laboratory perspective (Principles and Practice). (pp 131-156) Springer-Verlag Berlin Heidelberg New York.
31. **EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, UK [Internet].** EMBL-EBI 2015 [cited 2015 December 03]. Available from:<https://www.ebi.ac.uk/>

32. **Dawkins R**, Leelayuwat C, Gaudieri S, Tay G, Hui J, Cattley S, Martinez P, Kulski J. Genomics of the major histocompatibility complex: haplotypes, duplication, retroviruses and disease. *Immunol Rev.* 1999;167:275-304.
33. **Horton R**, Gibson R, Coghill P, Miretti M, Allcock RJ, Almeida J, Forbes S, Gilbert JG, Halls K, Harrow JL, Hart E, Howe K, Jackson DK, Palmer S, Roberts AN, Sims S, Stewart CA, Traherne JA, Trevanion S, Wilming L, Rogers J, de Jong PJ, Elliott JF, Sawcer S, Todd JA, Trowsdale J, Beck S. Variation analysis and gene annotation of eight MHC haplotypes: the MHC Haplotype Project. *Immunogenetics.* 2008; 60(1):1-18.
34. **Miretti MM**, Walsh EC, Ke X, Delgado M, Griffiths M, Hunt S, Morrison J, Whittaker P, Lander ES, Cardon LR, Bentley DR, Rioux JD, Beck S, Deloukas P. A high-resolution linkage-disequilibrium map of the human major histocompatibility complex and first generation of tag single-nucleotide polymorphisms. *Am J Hum Genet.* 2005; 76(4):634-46.
35. **Kotsch K**, Blasczyk R. The noncoding regions of HLA-DRB uncover interlineage recombinations as a mechanism of HLA diversification. *J Immunol.* 2000;15;165(10):5664-70.
36. **Brown JH**, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, Wiley DC. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature.* 1993;1;364(6432):33-9.
37. **Jardetzky TS**, Brown JH, Gorga JC, Stern LJ, Urban RG, Chi YI, Stauffacher C, Strominger JL, Wiley DC. Three-dimensional structure of a human class II histocompatibility molecule complexed with superantigen. *Nature*;1994;368, 711–718.
38. **Lee KH**, Wucherpfennig KW, and Wiley DC. Structure of a human insulin peptide–HLA-DQ8 complex and susceptibility to type 1 diabetes. *Nat. Immunol.* 2001;2:501.
39. **Falk K**, Rotzschke O, Stevanovic S, Jung G, Rammensee HG. Pool sequencing of natural HLA-DR, DQ, and DP ligands reveals detailed peptide motifs, constraints of processing, and general rules. *Immunogenetics*; 1994;39:230.
40. **Díaz G**, Amicosante M, Jaraquemada D, Butler RH, Guillén MV, Sánchez M, Nombela C, Arroyo J. Int Immunol. Functional analysis of HLA-DP polymorphism: a crucial role for DPβ residues 9, 11, 35, 55, 56, 69 and 84-87 in T cell allorecognition and peptide binding. 2003;15(5):565-76.

41. **Hollenbach JA**, Madbouly A, Gragert L, Vierra-Green C, Flesch S, Spellman S, Begovich A, Noreen H, Trachtenberg E, Williams T, Yu N, Shaw B, Fleischhauer K, Fernandez-Vina M, Maier M. A combined DPA1~DPB1 amino acid epitope is the primary unit of selection on the HLA-DP heterodimer. *Immunogenetics*. 2012;64(8):559-69.
42. **Solberg OD**, Mack SJ, Lancaster AK, Single RM, Tsai Y, Sanchez-Mazas A, Thomson G. Balancing selection and heterogeneity across the classical human leukocyte antigen loci: a meta-analytic review of 497 population studies. *Hum Immunol*. 2008;69(7):443-64.
43. **Astorga GP**, Williams RC Jr. Altered reactivity in mixed lymphocyte culture of lymphocytes from patients with rheumatoid arthritis. *Arthritis Rheum*. 1969;12(6):547-54.
44. **Stastny P**: Mixed lymphocyte cultures in rheumatoid arthritis. *J Clin Invest* 1976; 57: 1148-1 157.
45. **Stastny P**: Association of the B-cell alloantigen DRw4 with rheumatoid arthritis. *N Engl J Med*;1978; 298:869-871.
46. **McMichael AJ**, Sasazuki T, McDevitt HO, Payne RO. Increased frequency of HLA-CW3 and HLA-Dw4 in rheumatoid arthritis. *Arthritis Rheum*. 1977; 20:1037-1041.
47. **Panayi GS**, Wooley PH, Batchelor JR. HLA-DRw4 and rheumatoid arthritis. *Lancet*. 1979;31;1(8118):730.
48. **Duquesnoy RJ**, Marrari M, Hackbarth S, Zeevi A. Serological and cellular definition of a new HLA-DR associated determinant, MCI, and its association with rheumatoid arthritis. *Hum Immunol*; 1984;10: 165-176.
49. **Legrand L**, Lathrop GM, Marcelli-Barge A, Dryll A, Bardin T, Debeyre N, Poirier JC, Schmid M, Ryckewaert A, Dausset J: HLA-DR genotype risks in seropositive rheumatoid arthritis. *Am J Hum Genet*; 1984; 36:69-699.
50. **Gregersen PK**, Silver J, Winchester RJ. The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. *Arthritis Rheum*. 1987;30(11):1205-13.

51. **Winchester RJ**, Gregersen PK. The molecular basis of susceptibility to rheumatoid arthritis: the conformational equivalence hypothesis. *Springer Semin Immunopathol.* 1988;10(2-3):119-39.
52. **Winchester R**. The molecular basis of susceptibility to rheumatoid arthritis. *Adv Immunol* 1994;56: 389–466.
53. **Mackie SL**, Taylor JC, Martin SG; YEAR Consortium; UKRAG Consortium, Wordsworth P, Steer S, Wilson AG, Worthington J, Emery P, Barrett JH, Morgan AW. A spectrum of susceptibility to rheumatoid arthritis within HLA-DRB1: stratification by autoantibody status in a large UK population. *Genes Immun.* 2012;13(2):120-8.
54. **Raychaudhuri S**, Sandor C, Stahl EA, Freudenberg J, Lee HS, Jia X, Alfredsson L, Padyukov L, Klareskog L, Worthington J, Siminovitch KA, Bae SC, Plenge RM, Gregersen PK, de Bakker PI. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat Genet.* 2012;29;44(3):291-6.
55. **Wordsworth P**, Pile KD, Buckely JD, Lanchbury JS, Ollier B, Lathrop M, Bell JI. HLA heterozygosity contributes to susceptibility to rheumatoid arthritis. *Am J Hum Genet* 1992;51:585–91.
56. **Zanelli E**, Gonzalez-Gay MA, David CS. Could HLA-DRB1 be the protective locus in rheumatoid arthritis? *Immunol Today* 1995;16:274–8.
57. **van der Helm-van Mil AH**, Huizinga TW, Schreuder GMT, Breedveld FC, de Vries RR, Toes RE. An independent role of protective HLA class II alleles in rheumatoid arthritis severity and susceptibility. *Arthritis Rheum* 2005;9:2637–44.
58. **de Vries N** Tijssen H, van Riel PL, van de Putte LB. Reshaping the shared epitope hypothesis: HLA-associated risk for rheumatoid arthritis is encoded by amino acid substitutions at positions 67–74 of the HLA-DRB1 molecule. *Arthritis Rheum* 2002;46:921–8.
59. **Mattey DL**, Dawes PT, Gonzalez-Gay MA, Garcia-Porrua C, Thomson W, Hajeer AH, Ollier WE. HLA-DRB1 alleles encoding an aspartic acid at position 70 protect against development of rheumatoid arthritis. *J Rheumatol* 2001;28:232–9.
60. **du Montcel ST**, Michou L, Petit-Teixeira E, Osorio J, Lemaire I, Lasbleiz S, Pierlot C, Quillet P, Bardin T, Prum B, Cornelis F, Clerget-Darpoux F New classification of HLA-DRB1 alleles supports

- the shared epitope hypothesis of rheumatoid arthritis susceptibility. *Arthritis Rheum.* 2005;52(4):1063-8.
61. **Michou L**, Croiseau P, Petit-Teixeira E, du Montcel ST, Lemaire I, Pierlot C, Osorio J, Frigui W, Lasbleiz S, Quillet P, Bardin T, Prum B, Clerget-Darpoux F, Cornélis F; European Consortium on Rheumatoid Arthritis Families. European consortium on rheumatoid arthritis families. Validation of the reshaped shared epitope HLA-DRB1 classification in rheumatoid arthritis. *Arthritis Res Ther* 2006;8:R79.
 62. **Barnetche T**, Constantin A, Cantagrel A, Cambon-Thomsen A, Gourraud PA. New classification of HLA-DRB1 alleles in rheumatoid arthritis susceptibility: a combined analysis of worldwide samples. *Arthritis Res Ther* 2008;10:R26.
 63. **Morgan AW**, Haroon-Rashid L, Martin SG, Gooi HC, Worthington J, Thomson W, Barrett JH, Emery P. The shared epitope hypothesis in rheumatoid arthritis: evaluation of alternative classification criteria in a large UK Caucasian cohort. *Arthritis Rheum* 2008;58:1275–83.
 64. **van Zeven D**, Hazes JM, Zwinderman AH, Cats A, Schreuder GM, D'Amato J, Breedveld FC. Association of HLA-DR4 with a more progressive disease course in patients with rheumatoid arthritis. Results of a followup study. *Arthritis Rheum* 1991;34: 822–30.
 65. **Terao C**, Yano K, Ikari K, Furu M, Yamakawa N, Yoshida S, Hashimoto M, Ito H, Fujii T, Ohmura K, Yurugi K, Miura Y, Maekawa T, Taniguchi A, Momohara S, Yamanaka H, Mimori T, Matsuda F. Main Contribution of DRB1*04:05 Among the Shared Epitope Alleles and Involvement of DRB1 Amino Acid Position 57 in Association With Joint Destruction in Anti-Citrullinated Protein Antibody-Positive Rheumatoid. *Arthritis Arthritis Rheumatol.* 2015;67(7):1744-50.
 66. **Turesson C**, Weyand CM, Matteson EL. Genetics of rheumatoid arthritis: is there a pattern predicting extraarticular manifestations? *Arthritis Rheum.* 2004; 15;51(5):853-63.
 67. **Furukawa H**, Oka S, Shimada K, Hashimoto A, Tohma S. Human leukocyte antigen polymorphisms and personalized medicine for rheumatoid arthritis. *J Hum Genet.* 2015;60(11):691-6.
 68. **Holoshitz J**. The quest for better understanding of HLA-disease association: scenes from a road less travelled by. *Discov. Med:* 2013;16, 93–101.

69. **Morgan AW**, Thomson W, Martin SG; Yorkshire Early Arthritis Register Consortium, Carter AM; UK Rheumatoid Arthritis Genetics Consortium, Erlich HA, Barton A, Hocking L, Reid DM, Harrison P, Wordsworth P, Steer S, Worthington J, Emery P, Wilson AG, Barrett JH. Reevaluation of the interaction between HLA-DRB1 shared epitope alleles, PTPN22, and smoking in determining susceptibility to autoantibody-positive and autoantibody-negative rheumatoid arthritis in a large UK Caucasian population. *Arthritis Rheum.* 2009;2565–2576.
70. **van de Stadt LA**, de Koning MH, van de Stadt RJ, Wolbink G, Dijkmans BA, Hamann D, van Schaardenburg D. Development of the anti-citrullinated protein antibody repertoire prior to the onset of rheumatoid arthritis. *Arthritis Rheum.* 2011;63(11):3226-33.
71. **Kokkonen H**, Brink M, Hansson M, Lassen E, Mathsson-Alm L, Holmdahl R, Rönnelid J, Klareskog L, Rantapää-Dahlqvist. Associations of antibodies against citrullinated peptides with human leukocyte antigen-shared epitope and smoking prior to the development of rheumatoid arthritis. *Arthritis Res Ther.* 2015;20;17:125.
72. **Meyer O**, Nicaise-Roland P, Santos MD, Labarre C, Dougados M, Goupille P, Cantagrel A, Sibilia J, Combe B. Serial determination of cyclic citrullinated peptide autoantibodies predicted five-year radiological outcomes in a prospective cohort of patients with early rheumatoid arthritis. *Arthritis Res Ther.* 2006; 8(2):R40.
73. **Vander Cruyssen B**, Peene I, Cantaert T, Hoffman IE, De Rycke L, Veys EM, De Keyser F. Anti-citrullinated protein/peptide antibodies (ACPA) in rheumatoid arthritis: specificity and relation with rheumatoid factor. *Autoimmun Rev* 2005;4:468–74.
74. **van Gaalen FA**, van Aken J, Huizinga TW, Schreuder GM, Breedveld FC, Zanelli E, van Venrooij WJ, Verweij CL, Toes RE, de Vries RR. Association between HLA class II genes and autoantibodies to cyclic citrullinated peptides (CCPs) influences the severity of rheumatoid arthritis. *Arthritis Rheum.* 2004; 50(7):2113-21.
75. **van der Helm-van Mil AH**, Verpoort KN, Breedveld FC, Huizinga TW, Toes RE, de Vries RR. The HLA-DRB1 shared epitope alleles are primarily a risk factor for anti-cyclic citrullinated peptide antibodies and are not an independent risk factor for development of rheumatoid arthritis. *Arthritis Rheum.* 2006;54, 1117–1121.
76. **Verpoort KN**, Cheung K, Ioan-Facsinay A, van der Helm-van Mil AH, de Vries-Bouwstra JK, Allaart CF, Drijfhout JW, de Vries RR, Breedveld FC, Huizinga TW, Puijn GJ, Toes RE. Fine specificity of

the anti-citrullinated protein antibody response is influenced by the shared epitope alleles. *Arthritis Rheum.* 2007;56, 3949–3952.

77. **van der Woude D**, Houwing-Duistermaat JJ, Toes RE, Huizinga TW, Thomson W, Worthington J, van der Helm-van Mil AH, de Vries RR. Quantitative heritability of anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis. *Arthritis Rheum.* 60, 2009;916–923.
78. **Terao C**, Ohmura K, Ikari K, Kochi Y, Maruya E, Katayama M, Yurugi K, Shimada K, Murasawa A, Honjo S, Takasugi K, Matsuo K, Tajima K, Suzuki A, Yamamoto K, Momohara S, Yamanaka H, Yamada R, Saji H, Matsuda F, Mimori T. ACPA-negative RA consists of two genetically distinct subsets based on RF positivity in Japanese. *PLoS One.* 2012;7(7):e40067.
79. **Kroot EJ**, de Jong BA, van Leeuwen MA, Swinkels H, van den Hoogen FH, van't Hof M, van de Putte LB, van Rijswijk MH, van Venrooij WJ, van Riel PL The prognostic value of anti-cyclic citrullinated peptide antibody in patients with recent-onset rheumatoid arthritis. *Arthritis Rheum.* 2000;43, 1831–1835.
80. **Szodoray P**, Szabó Z, Kapitány A, Gyetvai A, Lakos G, Szántó S, Szücs G, Szekanecz Z. Anti-citrullinated protein/peptide autoantibodies in association with genetic and environmental factors as indicators of disease outcome in rheumatoid arthritis. *Autoimmun Rev.* 2010; 9(3):140-3.
81. **Hammer J**, Gallazzi F, Bono E, Karr RW, Guenot J, Valsasnini P, Nagy ZA, Sinigaglia F Peptide binding specificity of HLA-DR4 molecules: correlation with rheumatoid arthritis association. *J Exp Med.* 1995;1;181(5):1847-55.
82. **Roudier J**. Association of MHC and rheumatoid arthritis. Association of RA with HLA-DR4: the role of repertoire selection. *Arthritis Res*; 2000. 2, 217–220.
83. **Hill JA**, Southwood S, Sette A, Jevnikar AM, Bell DA, Cairns E Cutting edge: the conversion of arginine to citrulline allows for a high-affinity peptide interaction with the rheumatoid arthritis-associated HLA-DRB1*0401 MHC class II molecule. *J Immunol.* 2003;15; 171(2):538-41.
84. **Snir O**, Rieck M, Gebe JA, Yue BB, Rawlings CA, Nepom G, Malmström V, Buckner JH. Identification and functional characterization of T cells reactive to citrullinated vimentin in HLA-

- DRB1*0401-positive humanized mice and rheumatoid arthritis patients. *Arthritis Rheum*. 2011; 63(10):2873-83.
85. **Scally SW**, Petersen J, Law SC, Dudek NL, Nel HJ, Loh KL, Wijeyewickrema LC, Eckle SB, van Heemst J, Pike RN, McCluskey J, Toes RE, La Gruta NL, Purcell AW, Reid HH, Thomas R, Rossjohn J. A molecular basis for the association of the HLA-DRB1 locus, citrullination, and rheumatoid arthritis. *J Exp Med*. 2013;18;210(12):2569-82.
 86. **Miyadera H**, Tokunaga K. Associations of human leukocyte antigens with autoimmune diseases: challenges in identifying the mechanism. *J Hum Genet*. 2015;60(11):697-702.
 87. **Cosway E**, Anderson G, Garside P, Prendergast C. The thymus and rheumatology: should we care? *Curr Opin Rheumatol*. 2016;28(2):189-95.
 88. **Law SC**, Street S, Yu CH, Capini C, Ramnøruth S, Nel HJ, van Gorp E, Hyde C, Lau K, Pahau H, Purcell AW, Thomas R. T-cell autoreactivity to citrullinated autoantigenic peptides in rheumatoid arthritis patients carrying HLA-DRB1 shared epitope alleles. *Arthritis Res Ther*. 2012;17;14(3):R118).
 89. **Ling S**, Cline EN, Haug TS, Fox DA, Holoshitz J. Citrullinated calreticulin potentiates rheumatoid arthritis shared epitope signaling. *Arthritis Rheum* 2013;65(3):618–26.
 90. **Costenbader KH**, Prescott J, Zee RY, De Vivo I. Immunosenescence and rheumatoid arthritis: does telomere shortening predict impending disease? *Autoimmun Rev*; 2011;10(9):569–73.
 91. **LeRoy EC**, Black C, Fleischmajer R, Jablonska S, Krieg T, Medsger TA Jr, Rowell N, Wollheim F. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. *J Rheumatol*. 1988 Feb;15(2):202-5.
 92. **Chiffot H**, Fautrel B, Sordet C, Chatelus E, Sibilia J. Incidence and prevalence of systemic sclerosis: a systematic literature review. *Semin Arthritis Rheum* 2008; 37(4): 223-35.
 93. **Mayes MD**. Scleroderma epidemiology. *Rheum Dis Clin North A* 2003; 29(2): 239-54.
 94. **Rubio-Rivas M**, Royo C, Simeón CP, Corbella X, Fonollosa V. Mortality and survival in systemic sclerosis: systematic review and meta-analysis. *Semin Arthritis Rheum*. 2014;44(2):208-19.

95. **Tyndall AJ**, Bannert B, Vonk M, Airò P, Cozzi F, Carreira PE, Bancel DF, Allanore Y, Müller-Ladner U, Distler O, Iannone F, Pellerito R, Pilecky M, Miniati I, Ananieva L, Gurman AB, Damjanov N, Mueller A, Valentini G, Riemekasten G, Tikly M, Hummers L, Henriques MJ, Caramaschi P, Scheja A, Rozman B, Ton E, Kumánovics G, Coleiro B, Feierl E, Szucs G, Von Mühlen CA, Riccieri V, Novak S, Chizzolini C, Kotulska A, Denton C, Coelho PC, Kötter I, Simsek I, de la Pena Lefebvre PG, Hachulla E, Seibold JR, Rednic S, Stork J, Morovic-Vergles J, Walker UA. Causes and risk factors for death in systemic sclerosis: a study from the EULAR Scleroderma Trials and Research (EUSTAR) database. *Ann Rheum Dis*. 2010;69(10):1809-15.
96. **Simeón-Aznar CP**, Fonollosa-Plá V, Tolosa-Vilella C, Espinosa-Garriga G, Campillo-Grau M, Ramos-Casals M, García-Hernández FJ, Castillo-Palma MJ, Sánchez-Román J, Callejas-Rubio JL, Ortego-Centeno N, Egurbide-Arberas MV, Trapiella-Martínez L, Caminal-Montero L, Sáez-Comet L, Velilla-Marco J, Camps-García MT, de Ramón-Garrido E, Esteban-Marcos EM, Pallarés-Ferreres L, Navarrete-Navarrete N, Vargas-Hitos JA, Gómez de la Torre R, Salvador-Cervello G, Rios-Blanco JJ, Vilardell-Tarrés M; Spanish Scleroderma Study Group (SSSG); Autoimmune Diseases Study Group (GEAS); Spanish Society of Internal Medicine (SEMI). Registry of the Spanish Network for Systemic Sclerosis: Survival, Prognostic Factors, and Causes of Death. *Medicine (Baltimore)*. 2015;94(43):e1728
97. **Pattanaik D**, Brown M, Postlethwaite BC, Postlethwaite AE. Pathogenesis of Systemic Sclerosis. *Front Immunol*. 2015;8;6:272.
98. **Stern EP**, Denton CP. The Pathogenesis of Systemic Sclerosis. *Rheum Dis Clin North Am*. 2015;41(3):367-82.
99. **Cottrell TR**, Wise RA, Wigley FM, Boin F. The degree of skin involvement identifies distinct lung disease outcomes and survival in systemic sclerosis. *Ann Rheum Dis*. 2014;73(6).
100. **Ho YY**, Lagares D, Tager AM, Kapoor M. Fibrosis--a lethal component of systemic sclerosis. *Nat Rev Rheumatol*. 2014;10(7):390-402.
101. **Bhattacharyya S**, Wei J, Varga J. Understanding fibrosis in systemic sclerosis: shifting paradigms, emerging opportunities. *Nat. Rev. Rheumatol*;2012;8,42–54.
102. **Castelino FV**, Varga J. Emerging cellular and molecular targets in fibrosis: implications for scleroderma pathogenesis and targeted therapy. *Curr Opin Rheumatol*. 2014;26(6):607-14

103. **Varga J**, Abraham D. Systemic sclerosis: a prototypic multisystem fibrotic disorder. *J Clin Invest.* 2007;117(3):557-67.
104. **Kahaleh B**. Vascular disease in scleroderma: mechanisms of vascular injury. *Rheum Dis Clin North Am*;2008;34(1):57–71.
105. **Hasegawa M**, Takehara K. Potential immunologic targets for treating fibrosis in systemic sclerosis: a review focused on leukocytes and cytokines. *Semin Arthritis Rheum.* 2012;42(3):281-96.
106. **Manetti M**, Guiducci S, Romano E, Rosa I, Ceccarelli C, Mello T, et al. Differential expression of junctional adhesion molecules in different stages of systemic sclerosis. *Arthritis Rheum* (2013) 65(1):247–57.
107. **Lafyatis R**. Transforming growth factor β --at the centre of systemic sclerosis. *Nat Rev Rheumatol.* 2014;10(12):706-19.
108. **Maring JA**, Trojanowska M, ten Dijke P. Role of endoglin in fibrosis and scleroderma. *Int Rev Cell Mol Biol*; 2012;297:295–308.
109. **Postlethwaite AE**, Chiang TM. Platelet contributions to the pathogenesis of systemic sclerosis. *Curr Opin Rheumatol*;2007;19(6):574–9
110. **O'Reilly S**. Role of interleukin-13 in fibrosis, particularly systemic sclerosis. *Biofactors* 2013;39(6):593-6.
111. **Huang XL**, Wang YJ, Yan JW, Wan YN, Chen B, Li BZ, Yang GJ, Wang J. Role of anti-inflammatory cytokines IL-4 and IL-13 in systemic sclerosis. *Inflamm Res.* 2015;64(3-4):151-9.
112. **Desallais L**, Avouac J, Fréchet M, Elhai M, Ratsimandresy R, Montes M, Mouhsine H, Do H, Zagury JF, Allanore Y. Targeting IL-6 by both passive or active immunization strategies prevents bleomycin-induced skin fibrosis *Arthritis Res Ther.* 2014;24;16(4):R157
113. **van Bon L**, Affandi AJ, Broen J, Christmann RB, Marijnissen RJ, Stawski L, Farina GA, Stifano G, Mathes AL, Cossu M, York M, Collins C, Wenink M, Huijbens R, Hesselstrand R, Saxne T, DiMarzio M, Wuttge D, Agarwal SK, Reveille JD, Assassi S, Mayes M, Deng Y, Drenth JP, de Graaf J, den Heijer M, Kallenberg CG, Bijl M, Loof A, van den Berg WB, Joosten LA, Smith V, de Keyser F, Scorza

- R, Lunardi C, van Riel PL, Vonk M, van Heerde W, Meller S, Homey B, Beretta L, Roest M, Trojanowska M, Lafyatis R, Radstake TR. Proteome-wide analysis and CXCL4 as a biomarker in systemic sclerosis. *N Engl J Med*. 2014;370(5):433-43.
114. **Bandinelli F**, Del Rosso A, Gabrielli A, et al. CCL2, CCL3 and CCL5 chemokines in systemic sclerosis: the correlation with SSc clinical features and the effect of prostaglandin E1 treatment. *Clin Exp Rheumatol*;2012; 30:S44–S49.
 115. **Yamamoto T**, Eckes B, Krieg T. Effect of interleukin-10 on the gene expression of type I collagen, fibronectin, and decorin in human skinfibroblasts: differential regulation by transforming growth factor-beta and monocyte chemoattractant protein-1. *Biochem Biophys Res Commun*. 2001;6;281(1):200-5.
 116. **Shea BS**, Tager AM. Role of the lysophospholipid mediators lysophosphatidic acid and sphingosine 1-phosphate in lung fibrosis. *Proc Am Thorac Soc*, 2012; 9:102–110.
 117. **Wei J**, Fang F, Lam AP, Sargent JL, Hamburg E, Hinchcliff ME, Gottardi CJ, Atit R, Whitfield ML, Varga J. Wnt/beta-catenin signaling is hyperactivated in systemic sclerosis and induces Smad-dependent fibrotic responses in mesenchymal cells. *Arthritis Rheum* 2012; 64:2734–2745.
 118. **Dees C**, Tomcik M, Zerr P, Akhmetshina A, Horn A, Palumbo K, Beyer C, Zwerina J, Distler O, Schett G, Distler JH. Notch signalling regulates fibroblast activation and collagen release in systemic sclerosis. *Ann Rheum Dis*. 2011;70(7):1304-10.
 119. **Avouac J**, Clemessy M, Distler JH, Gasc JM, Ruiz B, Vacher-Lavenu MC, Wipff J, Kahan A, Boileau C, Corvol P, Allanore Y. Enhanced expression of ephrins and thrombospondins in the dermis of patients with early diffuse systemic sclerosis: potential contribution to perturbed angiogenesis and fibrosis. *Rheumatology (Oxford)* 2011; 50:1494–1504.
 120. **Gerber EE**, Gallo EM, Fontana SC, Davis EC, Wigley FM, Huso DL, Dietz HC. Integrin-modulating therapy prevents fibrosis and autoimmunity in mouse models of scleroderma. *Nature* 2014; 503:126–130.
 121. **O'Reilly S**. Innate immunity in systemic sclerosis pathogenesis. *Clin Sci (Lond)*. 2014;126(5):329-37.
 122. **Wu M**, Assassi S. The role of type 1 interferon in systemic sclerosis. *Front Immunol*. 2013;6;4:266.

123. **Ciechomska M**, Huigens CA, Hügler T, et al. Toll-like receptor-mediated, enhanced production of profibrotic TIMP-1 in monocytes from patients with systemic sclerosis: role of serum factors. *Ann Rheum Dis*. 2013;72(8):1382-9.
124. **Manetti M**. Deciphering the alternatively activated (M2) phenotype of macrophages in scleroderma. *Exp Dermatol*. 2015;24(8):576-8.
125. **Mosser DM**, Edwards JP. Exploring the full spectrum of macrophage activation. *Nat Rev Immunol*. 2008;8(12):958–69
126. **Farina GA**, York MR, Di Marzio M, Collins CA, Meller S, Homey B, Rifkin IR, Marshak-Rothstein A, Radstake TR, Lafyatis R. Poly(I:C) drives type I IFN- and TGFβ-mediated inflammation and dermal fibrosis simulating altered gene expression in systemic sclerosis. *J Invest Dermatol*. 2010;130(11):2583-93.
127. **Yao X**, Huang J, Zhong H, Shen N, Faggioni R, Fung M, Yao Y *Pharmacol Ther*. Targeting interleukin-6 in inflammatory autoimmune diseases and cancers. 2014;141(2):125-39.
128. **Rose-John S**. The soluble interleukin-6 receptor and related proteins. *Best Pract Res Clin Endocrinol Metab*. 2015;29(5):787-97.
129. **Muangchant C**, Pope JE. The significance of interleukin-6 and C-reactive protein in systemic sclerosis: a systematic literature review. *Clin Exp Rheumatol*. 2013;31(2 Suppl 76):122-34.
130. **Distler O**, Cozzio A. Systemic sclerosis and localized scleroderma--current concepts and novel targets for therapy. *Semin Immunopathol*. 2016;38(1):87-95.
131. **Hirahara K**, Nakayama T. CD4+ T-cell subsets in inflammatory diseases: beyond the Th1/Th2 paradigm. *Int Immunol*. 2016;28(4):163-71.
132. **Wynn TA** Fibrotic disease and the T(H)1/T(H)2 paradigm *Nat Rev Immunol*, 2004;4(8) 583–594.
133. **Bhogal RK**, Bona CA. Regulatory effect of extracellular signal-regulated kinases (ERK) on type I collagen synthesis in human dermal fibroblasts stimulated by IL-4 and IL-13. *Int. Rev. Immunol*. 2008;27, 472–496.

134. **Wen FQ**, Kohyama T, Liu X, Zhu YK, Wang H, Kim HJ, Kobayashi T, Abe S, Spurzem JR, Rennard SI. Interleukin-4- and interleukin-13-enhanced transforming growth factor-beta2 production in cultured human bronchial epithelial cells is attenuated by interferon-gamma. *Am J Respir Cell Mol Biol*. 2002;26(4):484-90.
135. **Fuschiotti P**. Role of IL-13 in systemic sclerosis. *Cytokine* 2011;56, 544–549.
136. **Patel DD**, Kuchroo VK. Cell Pathway in Human Immunity: Lessons from Genetics and Therapeutic Interventions. *Immunity*. 2015;15;43(6):1040-51.
137. **Yang X**, Yang J, Xing X, Wan L, Li M. Increased frequency of Th17 cells in systemic sclerosis is related to disease activity and collagen overproduction. *Arthritis Res Ther*. 2014;7;16(1):R4.
138. **Murata M**, Fujimoto M, Matsushita T, Hamaguchi Y, Hasegawa M, Takehara K, Komura K, Sato S. Clinical association of serum interleukin-17 levels in systemic sclerosis: is systemic sclerosis a Th17 disease? *J Dermatol Sci*. 2008;50(3):240-2.
139. **Okamoto Y**, Hasegawa M, Matsushita T, Hamaguchi Y, Huu DL, Iwakura Y, Fujimoto M, Takehara K. Potential roles of interleukin-17A in the development of skin fibrosis in mice. *Arthritis Rheum*. 2012;64(11):3726-35.
140. **Brembilla NC**, Montanari E, Truchetet ME, Raschi E, Meroni P, Chizzolini C. Th17 cells favor inflammatory responses while inhibiting type I collagen deposition by dermal fibroblasts: differential effects in healthy and systemic sclerosis fibroblasts. *Arthritis Res Ther*. 2013;10;15(5):R151.
141. **Higashi-Kuwata N**, Jinnin M, Makino T, Fukushima S, Inoue Y, Muchemwa FC, Yonemura Y, Komohara Y, Takeya M, Mitsuya H, Ihn H. Characterization of monocyte/ macrophage subsets in the skin and peripheral blood derived from patients with systemic sclerosis. *Arthritis Res Ther*. 2010;12(4):R128.
142. **Yukawa S**, Yamaoka K, Sawamukai N, Shimajiri S, Kubo S, Miyagawa I, Sonomoto K, Saito K, Tanaka Y. Dermal mast cell density in fingers reflects severity of skin sclerosis in systemic sclerosis. *Mod Rheumatol*. 2013;23(6):1151-7.

143. **Matsushita T**, Hasegawa M, Hamaguchi Y, Takehara K, Sato S. Longitudinal analysis of serum cytokine concentrations in systemic sclerosis: association of interleukin 12 elevation with spontaneous regression of skin sclerosis. *J Rheumatol*;2006;33:275–284.
144. **Asano Y**, Sato S. Vasculopathy in scleroderma. *Semin Immunopathol.* 2015;37(5):489-500.
145. **Yoshizaki A**, Yanaba K, Iwata Y, Komura K, Ogawa A, Akiyama Y, Muroi E, Hara T, Ogawa F, Takenaka M, Shimizu K, Hasegawa M, Fujimoto M, Tedder TF, Sato S. Cell adhesion molecules regulate fibrotic process via Th1/Th2/Th17 cell balance in a bleomycin-induced scleroderma model. *J Immunol.* 2010;15;185(4):2502-15.
146. **Antiga E**, Quaglino P, Bellandi S, Volpi W, Del Bianco E, Comessatti A, Osella-Abate S, De Simone C, Marzano A, Bernengo MG, Fabbri P, Caproni M. Regulatory T cells in the skin lesions and blood of patients with systemic sclerosis and morphea. *Br J Dermatol.* 2010;162(5):1056-63.
147. **Klein S**, Kretz CC, Ruland V, Stumpf C, Haust M, Hartschuh W, Hartmann M, Enk A, Suri-Payer E, Oberle N, Krammer PH, Kuhn A. Reduction of regulatory T cells in skin lesions but not in peripheral blood of patients with systemic scleroderma *Ann Rheum Dis.* 2011;70(8):1475-81.
148. **MacDonald KG**, Dawson NA, Huang Q, Dunne JV, Levings MK, Broady R Regulatory T cells produce profibrotic cytokines in the skin of patients with systemic sclerosis. *J Allergy Clin Immunol.* 2015;135(4):946-e9.
149. **Sakkas LI**, Bogdanos DP. Systemic sclerosis: New evidence re-enforces the role of B cells. *Autoimmun Rev.* 2016;15(2):155-61.
150. **Baroni SS**, Santillo M, Bevilacqua F, Luchetti M, Spadoni T, Mancini M, Fraticelli P, Sambo P, Funaro A, Kazlauskas A, Avvedimento EV, Gabrielli A. Stimulatory autoantibodies to the PDGF receptor in systemic sclerosis. *N Engl J Med.* 2006; 22;354(25):2667-76.
151. **Sato S**, Hayakawa I, Hasegawa M, Fujimoto M, Takehara K. Function blocking autoantibodies against matrix metalloproteinase-1 in patients with systemic sclerosis. *J Invest Dermatol.* 2003;120(4):542-7.
152. **Nishijima C**, Hayakawa I, Matsushita T, Komura K, Hasegawa M, Takehara K, Sato S. Autoantibody against matrix metalloproteinase-3 in patients with systemic sclerosis. *Clin Exp Immunol.* 2004;138(2):357-63.

153. **Svegliati S**, Olivieri A, Campelli N, Luchetti M, Poloni A, Trappolini S, Moroncini G, Bacigalupo A, Leoni P, Avvedimento EV, Gabrielli A. Stimulatory autoantibodies to PDGF receptor in patients with extensive chronic graft-versus-host disease. *Blood*. 2007;1;110(1):237-41.
154. **Maddur MS**, Sharma M, Hegde P, Stephen-Victor E, Pulendran B, Kaveri SV, Bayry J.B cells Human B cells induce dendritic cell maturation and favour Th2 polarization by inducing OX-40 ligand. *Nat Commun*. 2014;9;5:4092.
155. **Gilbane AJ**, Denton CP, Holmes AM. Scleroderma pathogenesis: a pivotal role for fibroblasts as effector cells. *Arthritis Res Ther*. 2013;15(3):215.
156. **Hinz B**, Dugina V, Ballestrem C, Wehrle-Haller B, Chaponnier C. Alpha-smooth muscle actin is crucial for focal adhesion maturation in myofibroblasts. *Mol Biol Cell*. 2003;14(6):2508-19.
157. **Piera-Velazquez S**, Mendoza FA, Jimenez SA. Endothelial to Mesenchymal Transition (EndoMT) in the Pathogenesis of Human Fibrotic Diseases. *J Clin Med*. 2016;11;5(4).
158. **Poslethwaite AE**, Shigemitsu H, Kanagat S. Cellular origins of fibroblasts: Possible implications for organ fibrosis in systemic sclerosis. *Curr. Opin. Rheumatol*. 2004, 16, 733–738.
159. **Nikitorowicz-Buniak J**, Denton CP, Abraham D, Stratton R. Partially Evoked Epithelial-Mesenchymal Transition (EMT) Is Associated with Increased TGF β Signaling within Lesional Scleroderma Skin. *PLoS One*. 2015;28;10(7).
160. **Rajkumar VS**, Howell K, Csiszar K, Denton CP, Black CM, Abraham DJ Shared expression of phenotypic markers in systemic sclerosis indicates a convergence of pericytes and fibroblasts to a myofibroblast lineage in fibrosis. *Arthritis Res Ther*. 2005;7(5):R1113-23.
161. **Nicolosi PA**, Tombetti E, Maugeri N, Rovere-Querini P, Brunelli S, Manfredi AA. Vascular Remodelling and Mesenchymal Transition in Systemic Sclerosis. *Stem Cells Int*. 2016;2016:4636859.
162. **Shi-wen X**, Thompson K, Khan K, Liu S, Murphy-Marshman H, Baron M, Denton CP, Leask A, Abraham DJ Focal adhesion kinase and reactive oxygen species contribute to the persistent fibrotic phenotype of lesional scleroderma fibroblasts. *Rheumatology (Oxford)*. 2012; 51(12):2146-54.

163. **Budd DC**, Holmes AM. Targeting TGF β superfamily ligand accessory proteins as novel therapeutics for chronic lung disorders. *Pharmacol Ther.* 2012; 135(3):279-91.
164. **Cipriani P**, Di Benedetto P, Ruscitti P, Capece D, Zazzeroni F, Liakouli V, Pantano I, Berardicurti O, Carubbi F, Pecetti G, Turricchia S, Alesse E, Iglarz M, Giacomelli R. The Endothelial-mesenchymal Transition in Systemic Sclerosis Is Induced by Endothelin-1 and Transforming Growth Factor- β and May Be Blocked by Macitentan, a Dual Endothelin-1 Receptor Antagonist. *J Rheumatol.* 2015;42(10):1808-16.
165. **Parker MW**, Rossi D, Peterson M, Smith K, Sikström K, White ES, Connett JE, Henke CA, Larsson O, Bitterman PB. Fibrotic extracellular matrix activates a profibrotic positive feedback loop. *J Clin Invest* 2014; 124:1622–1635.
166. **Maeda T**, Sakabe T, Sunaga A, Sakai K, Rivera AL, Keene DR, Sasaki T, Stavnezer E, Iannotti J, Schweitzer R, Ilic D, Baskaran H, Sakai T. Conversion of mechanical force into TGF-beta-mediated biochemical signals. *Curr Biol* 2011; 21:933–941.
167. **Reich A**, Meurer M, Eckes B, Friedrichs J, Muller D.J. Surface morphology and mechanical properties of fibroblasts from scleroderma patients. *J. Cell. Mol. Med.* 2009;13, 1644–1652.
168. **Martin PJ**, Counts GW Jr, Appelbaum FR, Lee SJ, Sanders JE, Deeg HJ, Flowers ME, Syrjala KL, Hansen JA, Storb RF, Storer BE. Life expectancy in patients surviving more than 5 years after hematopoietic cell transplantation. *J Clin Oncol.* 2010;28(6):1011-6.
169. **Wingard JR**, Majhail NS, Brazauskas R, Wang Z, Sobocinski KA, Jacobsohn D, Sorrow ML, Horowitz MM, Bolwell B, Rizzo JD, Socié G. Long-term survival and late deaths after allogeneic hematopoietic cell transplantation. *J Clin Oncol.* 2011;29(16):2230-9.
170. **Filipovich AH**, Weisdorf D, Pavletic S, Socie G, Wingard JR, Lee SJ, Martin P, Chien J, Przepiorka D, Couriel D, Cowen EW, Dinndorf P, Farrell A, Hartzman R, Henslee-Downey J, Jacobsohn D, McDonald G, Mittleman B, Rizzo JD, Robinson M, Schubert M, Schultz K, Shulman H, Turner M, Vogelsang G, Flowers ME. 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(12):945-56.

171. **Martires KJ**, Baird K, Steinberg SM, Grkovic L, Joe GO, Williams KM, Mitchell SA, Datiles M, Hakim FT, Pavletic SZ, Cowen EW. Sclerotic type chronic GVHD of the skin: clinical risk factors, laboratory markers, and burden of disease. *Blood* 2011; 118: 4250–4257.
172. **Canady J**, Karrer S, Fleck M, Bosserhoff AK. Fibrosing connective tissue disorders of the skin: molecular similarities and distinctions. *J Dermatol Sci.* 2013;70(3):151-8.
173. **Inamoto Y**, Storer BE, Petersdorf EW, Nelson JL, Lee SJ, Carpenter PA, Sandmaier BM, Hansen JA, Martin PJ, Flowers ME. Incidence, risk factors and outcomes of sclerosis in patients with chronic graft-versus host disease. *Blood.* 2013;20;121(25):5098-103.
174. **Uhm J**, Hamad N, Shin EM, Michelis FV, Shanavas M, Gupta V, Kuruvilla J, Lipton JH, Messner HA, Seftel M, Kim DD. Incidence, risk factors, and long-term outcomes of sclerotic graft-versus-host disease after allogeneic hematopoietic cell transplantation. *Biol Blood Marrow Transplant.* 2014;20(11):1751-7.
175. **Detrait MY**, Morisset S, Peffault de Latour R, Yakoub-Agha I, Crocchiolo R, Tabrizi R, Bay JO, Chevalier P, Barraco F, Raus N, Vigouroux S, Magro L, Mohty M, Milpied N, Blaise D, Socié G, Michallet M. Pre-transplantation risk factors to develop sclerotic chronic GvHD after allogeneic HSCT: a multicenter retrospective study from the Société Française de Greffe de Moelle et de Thérapie Cellulaire (SFGM-TC). *Bone Marrow Transplant.* 2015;50(2):253-8.
176. **Shulman HM**, Kleiner D, Lee SJ, Morton T, Pavletic SZ, Farmer E, Moresi JM, Greenson J, Janin A, Martin PJ, McDonald G, Flowers ME, Turner M, Atkinson J, Lefkowitz J, Washington MK, Prieto VG, Kim SK, Argenyi Z, Diwan AH, Rashid A, Hiatt K, Couriel D, Schultz K, Hymes S, Vogelsang GB. 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.
177. **Jedlickova Z**, Burlakova I, Bug G, Baumann H, Schwerdtfeger R, Schleuning M. Therapy of sclerodermatous chronic graft-versus-host disease with mammalian target of rapamycin inhibitors. *Biol Blood Marrow Transplant.* 2011;17(5):657-63.
178. **Blazar BR**, Murphy WJ, Abedi M *Nat Rev Immunol.* 2012;11;12(6):443-58. Advances in graft-versus-host disease biology and therapy.
179. **Ferrara JL**, Levine JE, Reddy P, Holler E Graft-versus-host disease. *Lancet.* 2009, 2;373:1550-61.

180. **Min CK.** The pathophysiology of chronic graft-versus-host disease: the unveiling of an enigma. Korean J Hematol. 2011;46(2):80-7.
181. **Chu YW,** Gress RE. Murine models of chronic graft-versus-host disease: insights and unresolved issues. Biol Blood Marrow Transplant. 2008;14:365–378.
182. **Choi SW,** Levine JE, Ferrara JL. Pathogenesis and management of graft-versus-host disease. Immunol Allergy Clin North Am. 2010;30:75–101.
183. **Schroeder MA,** DiPersio JF. Mouse models of graft-versus- host disease: advances and limitations. Dis Model Mech. 2011;4(3):318-33 .
184. **Zhang Y,** Hexner E, Frank D, Emerson SG: CD4+ T cells generated de novo from donor hemopoietic stem cells mediate the evolution from acute to chronic graft-versus-host disease. J Immunol. 2007;179: 3305-3314.
185. **Dertschnig S,** Nusspaumer G, Ivanek R, Hauri-Hohl MM, Holländer GA, Krenger W. Epithelial cytoprotection sustains ectopic expression of tissue-restricted antigens in the thymus during murine acute GVHD. Blood. 2013;122(5):837–841.
186. **Bolton HA,** Zhu E, Terry AM, Guy TV, Koh WP, Tan SY, Power CA, Bertolino P, Lahl K, Sparwasser T, Shklovskaya E, Fazekas de St Groth B. Selective Treg reconstitution during lymphopenia normalizes DC costimulation and prevents graft-versus-host disease. J Clin Invest. 2015;125(9):3627–3641.
187. **Sarantopoulos S,** Blazar BR, Cutler C, Ritz J. B cells in chronic graft-versus-host disease. Biol Blood Marrow Transplant. 2015;21(1):16-23.
188. **Zhang C,** Todorov I, Zhang Z, et al. Donor CD4+ T and B cells in transplants induce chronic graft-versus-host disease with autoimmune manifestations. Blood. 2006;107:2993–3001.
189. **Lee YJ,** Min HS, Kang EH, Park HJ, Jeon YK, Kim JH, Wu HG, Lee EB, Park CG, Yoon SS, Park SH, Jung KC. Sclerodermatous chronic graft-versus-host disease induced by host T-cell-mediated autoimmunity. Immunol Cell Biol. 2012;90(3):358-67.
190. **Socié G,** Ritz J. Current issues in chronic graft-versus-host disease. Blood. 2014;17;124(3):374-84.

191. **Yeh SP**, Liao YM, Lo WJ, Lin CL, Bai LY, Lin CY, Hsieh CY, Chang YC, Huang YT, Chiu CF. Kinetics of T helper subsets and associated cytokines correlate well with the clinical activity of graft-versus-host disease. *PLoS One*. 2012;7(9):e44416.
192. **Wu Y**, Bastian D, Schutt S, Nguyen H, Fu J, Heinrichs J, Xia C, Yu XZ. Essential Role of Interleukin-12/23p40 in the Development of Graft-versus-Host Disease in Mice. *Biol Blood Marrow Transplant*. 2015;21(7):1195-204.
193. **Okamoto S**, Fujiwara H, Nishimori H, Matsuoka K, Fujii N, Kondo E, Tanaka T, Yoshimura A, Tanimoto M, Maeda Y. Anti-IL-12/23 p40 antibody attenuates experimental chronic graft-versus-host disease via suppression of IFN- γ /IL-17-producing cells. *J Immunol*. 2015;194(3):1357-63.
194. **Sprent J**, Kishimoto H. The thymus and central tolerance. *Transplantation* 2001;72(Suppl 8):S25-8.
195. **Sakoda Y**, Hashimoto D, Asakura S, Takeuchi K, Harada M, Tanimoto M and Teshima T. Donor-derived thymic-dependent T cells cause chronic graft-versus-host disease. *Blood*;2007;109:1756-1764.
196. **Na IK**, Lu SX, Yim NL, et al. The cytolytic molecules Fas ligand and TRAIL are required for murine thymic graft-versus-host disease. *J Clin Invest*. 2010;120(1):343–356. Wu T, Young JS, Johnston H, et al. Thymic damage, impaired negative selection, and development of chronic graft-versus-host disease caused by donor CD4⁺ and CD8⁺ T cells. *J Immunol*. 2013;191(1):488–499.
197. **Klein L**, Hinterberger M, Wirnsberger G, Kyewski B. Antigen presentation in the thymus for positive selection and central tolerance induction. *Nat Rev Immunol* 2009;9(12):833-844.
198. **Dertschnig S**, Hauri-Hohl MM, Vollmer M, Holländer GA, Krenger W. Impaired thymic expression of tissue-restricted antigens licenses the de novo generation of autoreactive CD4⁺ T cells in acute GVHD. *Blood*. 2015; 23;125(17):2720-3.
199. **Zorn E**, Kim HT, Lee SJ, Floyd BH, Litsa D, Arumugarajah S, Bellucci R, Alyea EP, Antin JH, Soiffer RJ, Ritz J. Reduced frequency of FOXP3⁺ CD4⁺CD25⁺ regulatory T cells in patients with chronic graft-versus-host disease. *Blood*. 2005;15;106(8):2903-11.

200. **Yuan X**, Cheng G, Malek TR. The importance of regulatory T-cell heterogeneity in maintaining self-tolerance. *Immunol Rev.* 2014;259(1):103–114.
201. **Alho AC**, Kim HT, Chammas MJ, Reynolds CG, Matos TR, Forcade E, Whangbo J, Nikiforow S, Cutler CS, Koreth J, Ho VT, Armand P, Antin JH, Alyea EP, Lacerda JF, Soiffer RJ, Ritz J. Unbalanced recovery of regulatory and effector T cells after allogeneic stem cell transplantation contributes to chronic GVHD. *Blood.* 2016 Feb 4;127(5):646-57.
202. **Anderson BE**, McNiff J, Yan J, Doyle H, Mamula M, Shlomchik MJ, Shlomchik WD. Memory CD4+ T cells do not induce graft-versus-host disease. *J Clin Invest.* 2003;112(1):101–108.
203. **Bleakley M**, Heimfeld S, Loeb KR, Jones LA, Chaney C, Seropian S, Gooley TA, Sommermeyer F, Riddell SR, Shlomchik WD. Outcomes of acute leukemia patients transplanted with naive T cell-depleted stem cell grafts. *J Clin Invest.* 2015;125(7):2677–2689.
204. **Miklos DB**, Kim HT, Miller KH, Guo L, Zorn E, Lee SJ, Hochberg EP, Wu CJ, Alyea EP, Cutler C, Ho V, Soiffer RJ, Antin JH, Ritz J. Antibody responses to H-Y minor histocompatibility antigens correlate with chronic graft-versus-host disease and disease remission. *Blood*;2005,105: 2973-2978.
205. **Nakasone H**, Tian L, Sahaf B, Kawase T, Schoenrock K, Perloff S, Ryan CE, Paul J, Popli R, Wu F, Otani JM, Collier J, Warren EH 3rd, Miklos DB. Allogeneic HY antibodies detected 3 months after female-to-male HCT predict chronic GVHD and nonrelapse mortality in humans. *Blood.* 2015;14;125(20):3193-201.
206. **Sahaf B**, Yang Y, Arai S, Herzenberg LA, Herzenberg LA, Miklos DB. H-Y antigen-binding B cells develop in male recipients of female hematopoietic cells and associate with chronic graft vs. host disease. *Proc Natl Acad Sci U S A.* 2013; 19;110(8):3005-10.
207. **Patriarca F**, Skert C, Sperotto A, Zaja F, Falletti E, Mestroni R, Kikic F, Calistri E, Fili C, Geromin A, Cerno M, Fanin R. The development of autoantibodies after allogeneic stem cell transplantation is related with chronic graft-vs-host disease and immune recovery. *Exp Hematol.* 2006 Mar;34(3):389-96.
208. **Sarantopoulos S**, Blazar BR, Cutler C, Ritz J. B cells in chronic graft-versus-host disease. *Biol Blood Marrow Transplant.* 2015;21(1):16-23.

209. **Jin H**, Ni X, Deng R, Song Q, Young J, Cassady K, Zhang M, Forman S, Martin PJ, Liu Q, Zeng D. Antibodies from donor B cells perpetuate cutaneous chronic graft-versus-host disease in mice. *Blood*. 2016;16. pii: blood-2015-09-668145. [Epub ahead of print]
210. **Young JS**, Wu T, Chen Y, Zhao D, Liu H, Yi T, Johnston H, Racine J, Li X, Wang A, Todorov I, Zeng D. Donor B cells in transplants augment clonal expansion and survival of pathogenic CD4+ T cells that mediate autoimmune-like chronic graft-versus-host disease. *J Immunol*. 2012;1;189(1):222-33.
211. **Sarantopoulos S**, Ritz J. Aberrant B-cell homeostasis in chronic GVHD. *Blood*. 2015;12;125(11):1703-7.
212. **Allen JL**, Tata PV, Fore MS, Wooten J, Rudra S, Deal AM, Sharf A, Hoffert T, Roehrs PA, Shea TC, Serody JS, Richards KL, Jagasia M, Lee SJ, Rizzieri D, Horwitz ME, Chao NJ, Sarantopoulos S. Increased BCR responsiveness in B cells from patients with chronic GVHD. *Blood*. 2014;27;123(13):2108-15.
213. **de Masson A**, Bouaziz JD, Le Buanec H, Robin M, O'Meara A, Parquet N, Rybojad M, Hau E, Monfort JB, Branchtein M, Michonneau D, Dessirier V, Sicre de Fontbrune F, Bergeron A, Itzykson R, Dhédin N, Bengoufa D, Peffault de Latour R, Xhaard A, Bagot M, Bensussan A, Socié G. CD24(hi)CD27(+) and plasmablast-like regulatory B cells in human chronic graft-versus-host disease. *Blood*. 2015;125(11):1830-1839.
214. **Le Huu D**, Matsushita T, Jin G, Hamaguchi Y, Hasegawa M, Takehara K, Tedder TF, Fujimoto M. Donor-derived regulatory B cells are important for suppression of murine sclerodermatous chronic graft-versus-host disease. *Blood*. 2013;121(16):3274-3283.
215. **Wynn TA**. Fibrotic disease and the T(H)1/T(H)2 paradigm. *Nat Rev Immunol*. 2004;4:583-594.
216. **Zhang Y**, McCormick LL, Desai SR, Wu C, Gilliam AC. Murine sclerodermatous graft-versus-host disease, a model for human scleroderma: cutaneous cytokines, chemokines, and immune cell activation. *J Immunol*. 2002; 168: 3088–3098.
217. **Zhou L**, Askew D, Wu C, Gilliam AC. Cutaneous gene expression by DNA microarray in murine sclerodermatous graft-versus-host disease, a model for human scleroderma. *J Invest Dermatol*. 2007;127(2):281-92.

218. **Koreth J**, Matsuoka K, Kim HT, McDonough SM, Bindra B, Alyea EP 3rd, Armand P, Cutler C, Ho VT, Treister NS, Bienfang DC, Prasad S, Tzachanis D, Joyce RM, Avigan DE, Antin JH, Ritz J, Soiffer RJ. Interleukin-2 and regulatory T cells in graft-versus-host disease. *N Engl J Med*. 2011;1;365(22):2055-66.
219. **Ma HH**, Ziegler J, Li C, Sepulveda A, Bedeir A, Grandis J, Lentzsch S, Mapara MY. Sequential activation of inflammatory signaling pathways during graft-versus-host disease (GVHD): early role for STAT1 and STAT3. *Cell Immunol*. 2011;268(1):37-46.
220. **Flynn R**, Paz K, Du J, Reichenbach DK, Taylor PA, Panoskaltsis-Mortari A, Vulic A, Luznik L, MacDonald KK, Hill GR, Nyuydzefe MS, Weiss JM, Chen W, Trzeciak A, Serody JS, Aguilar EG, Murphy WJ, Maillard I, Munn D, Koreth J, Cutler CS, Antin JH, Ritz J, Waksal SD, Zanin-Zhorov A, Blazar BR. Targeted Rho-associated kinase 2 inhibition suppresses murine and human chronic GVHD through a Stat3-dependent mechanism. *Blood*. 2016;28;127(17):2144-54.
221. **Hill GR**, Olver SD, Kuns RD, Varelías A, Raffelt NC, Don AL, Markey KA, Wilson YA, Smyth MJ, Iwakura Y, Tocker J, Clouston AD, Macdonald KP. Stem cell mobilization with G-CSF induces type 17 differentiation and promotes scleroderma. *Blood*. 2010;116(5):819–828.
222. **Normanton M**, Marti LC. Current data on IL-17 and Th17 cells and implications for graft versus host disease. *Einstein (Sao Paulo)*. 2013;11(2):237-46.
223. **Zhang Y**, McCormick LL, Gilliam AC. Latency associated peptide prevents skin fibrosis in murine sclerodermatous graft-versus-host disease, a model for human scleroderma. *J Invest Dermatol*. 2003;121(4):713-719.
224. **Banovic T**, MacDonald KP, Morris ES, Rowe V, Kuns R, Don A, Kelly J, Ledbetter S, Clouston AD, Hill GR. TGF-beta in allogeneic stem cell transplantation: friend or foe? *Blood*. 2005 Sep 15;106(6):2206-14.
225. **Chu YW**, Gress RE: Murine models of chronic graft-versus host disease: insights and unresolved issues. *Biol Blood Marrow Transplant*; 2008;14: 365-378.
226. **Pidala J**, Sarwal M, Roedder S, Lee SJ. Biologic markers of chronic GVHD. *Bone Marrow Transplant*. 2014;49(3):324-31.

227. **Rubtsov YP**, Rudensky AY: TGFbeta signalling in control of T-cell-mediated self-reactivity. *Nat Rev Immunol*; 2007;7: 443- 453.
228. **McCormick LL**, Zhang Y, Tootell E and Gilliam AC: Anti-TGF beta treatment prevents skin and lung fibrosis in murine sclerodermatous graft-versus-host disease: a model for human scleroderma. *J Immunol*;1999;163: 5693-5699.
229. **Dees C**, Beyer C, Distler A, Soare A, Zhang Y, Palumbo-Zerr K, Distler O, Schett G, Sandner P, Distler JH. Stimulators of soluble guanylate cyclase (sGC) inhibit experimental skin fibrosis of different aetiologies. *Ann Rheum Dis*. 2015;74(8):1621-5.
230. **Ruzek MC**, Jha S, Ledbetter S, Richards SM, Garman RD. A modified model of graft-versus-host-induced systemic sclerosis (scleroderma) exhibits all major aspects of the human disease. *Arthritis Rheum*. 2004;50(4):1319-31.
231. **Le Huu D**, Matsushita T, Jin G, Hamaguchi Y, Hasegawa M, Takehara K, Tedder TF, Fujimoto M. Donor-derived regulatory B cells are important for suppression of murine sclerodermatous chronic graft-versus-host disease. *Blood*. 2013;18;121(16):3274-83.
232. **Dobrota R**, Maurer B, Graf N, Jordan S, Mihai C, Kowal-Bielecka O, Allanore Y, Distler O; EUSTAR coauthors Prediction of improvement in skin fibrosis in diffuse cutaneous systemic sclerosis: a EUSTAR analysis. *Ann Rheum Dis*. 2016. pii: annrheumdis-2015-208024 (Epub ahead of print).
233. **Zerr P**, Distler A, Palumbo-Zerr K, Tomcik M, Vollath S, Dees C, Egberts F, Tinazzi I, Del Galdo F, Distler O, Schett G, Spriewald BM, Distler JH. Combined inhibition of c-Abl and PDGF receptors for prevention and treatment of murine sclerodermatous chronic graft-versus-host disease. *Am J Pathol*. 2012;181(5):1672-80.
234. **Rice LM**, Padilla CM, McLaughlin SR, Mathes A, Ziemek J, Goummih S, Nakerakanti S, York M, Farina G, Whitfield ML, Spiera RF, Christmann RB, Gordon JK, Weinberg J, Simms RW, Lafyatis R. Fresolimumab treatment decreases biomarkers and improves clinical symptoms in systemic sclerosis patients. *J Clin Invest*. 2015;1;125(7):2795-807.
235. **Fleming JN**, Shulman HM, Nash RA, Johnson PY, Wight TN, Gown A, Schwartz SM. Cutaneous chronic graft-versus-host disease does not have the abnormal endothelial phenotype or vascular rarefaction characteristic of systemic sclerosis. *PLoS One*. 2009;9;4(7):e6203.

236. **Greenblatt MB**, Sargent JL, Farina G, Tsang K, Lafyatis R, Glimcher LH, Whitfield ML, Aliprantis AO. Interspecies comparison of human and murine scleroderma reveals IL-13 and CCL2. *Am J Pathol*. 2012;180(3):1080-94.
237. **Sargent JL**, Li Z, Aliprantis AO, Greenblatt M, Lemaire R, Wu MH, Wei J, Taroni J, Harris A, Long KB, Burgwin C, Artlett CM, Blankenhorn EP, Lafyatis R, Varga J, Clark SH, Whitfield ML. Interspecies comparative genomics identifies optimal mouse models of systemic sclerosis. *Arthritis Rheumatol*. 2016. doi: 10.1002/art.39658. [Epub ahead of print]
238. **Milano A**, Pendergrass SA, Sargent JL, George LK, McCalmont TH, Connolly MK, Whitfield ML. Molecular subsets in the gene expression signatures of scleroderma skin. *PLoS One*. 2008;16;3(7):e2696.
239. **Pendergrass SA**, Lemaire R, Francis IP, Mahoney JM, Lafyatis R, Whitfield ML. Intrinsic gene expression subsets of diffuse cutaneous systemic sclerosis are stable in serial skin biopsies. *J Invest Dermatol*. 2012;132(5):1363-73.
240. **Hinchcliff M**, Huang CC, Wood TA, Matthew Mahoney J, Martyanov V, Bhattacharyya S, Tamaki Z, Lee J, Carns M, Podlasky S, Sirajuddin A, Shah SJ, Chang RW, Lafyatis R, Varga J, Whitfield ML. Molecular signatures in skin associated with clinical improvement during mycophenolate treatment in systemic sclerosis. *J Invest Dermatol*. 2013; 133(8):1979-89.
241. **Johnson ME**, Mahoney JM, Taroni J, Sargent JL, Marmarelis E, Wu MR, Varga J, Hinchcliff ME, Whitfield ML. Experimentally-derived fibroblast gene signatures identify molecular pathways associated with distinct subsets of systemic sclerosis patients in three independent cohorts. *PLoS One*. 2015;21;10(1).
242. **Kayser C**, Fritzler MJ. Autoantibodies in systemic sclerosis: unanswered questions. *Front Immunol*. 2015;5;6:167.
243. **Colaci M**, Giuggioli D, Manfredi A, Vacchi C, Della Casa G, Ferri C. Radiological thymus alterations in systemic sclerosis: our experience and a review of the literature. *Rheumatology (Oxford)*. 2014;53(4):732-6.
244. **Blagojevic J**, Matucci-Cerinic M. Systemic sclerosis: clinical manifestations. *Curr Rheumatol Rev*. 2010;6(4):295-304.

245. **Aschwanden M**, Halter JP, Walker UA, Staub D, Tichelli A, Daikeler T, Jaeger KA, Tyndall A. Nail fold capillaroscopy differs widely between systemic sclerosis and chronic graft vs host disease of the skin. *Rheumatology (Oxford)*. 2011;50(6):1168-9.
246. **Hofstee HM**, de Waal TT, Zweegman S, Voskuyl AE, Smulders YM, Schoordijk MC, Janssen JJ, Serné EH. Nailfold capillary abnormalities in sclerodermatous chronic GVHD. *Bone Marrow Transplant*. 2013;48(12):1574-7.
247. **Reveille JD**, Fischbach M, McNearney T, Friedman AW, Aguilar MB, Lisse J, Fritzler MJ, Ahn C, Arnett FC; GENISOS Study Group. Systemic sclerosis in 3 US ethnic groups: a comparison of clinical, sociodemographic, serologic, and immunogenetic determinants. *Semin Arthritis Rheum*. 2001;30(5):332-46.
248. **Gladman DD**, Kung TN, Siannis F, Pellett F, Farewell VT, Lee P. HLA markers for susceptibility and expression in scleroderma. *J Rheumatol* 2005;32(8):1481–7.
249. **Gorlova O**, Martin JE, Rueda B, Koeleman BP, Ying J, Teruel M, Diaz-Gallo LM, Broen JC, Vonk MC, Simeon CP, Alizadeh BZ, Coenen MJ, Voskuyl AE, Schuerwegh AJ, van Riel PL, Vanthuyne M, van 't Slot R, Italiaander A, Ophoff RA, Hunzelmann N, Fonollosa V, Ortego-Centeno N, González-Gay MA, García-Hernández FJ, González-Escribano MF, Airo P, van Laar J, Worthington J, Hesselstrand R, Smith V, de Keyser F, Houssiau F, Chee MM, Madhok R, Shiels PG, Westhovens R, Kreuter A, de Baere E, Witte T, Padyukov L, Nordin A, Scorza R, Lunardi C, Lie BA, Hoffmann-Vold AM, Palm O, García de la Peña P, Carreira P; Spanish Scleroderma Group, Varga J, Hinchcliff M, Lee AT, Gourh P, Amos CI, Wigley FM, Hummers LK, Nelson JL, Riemekasten G, Herrick A, Beretta L, Fonseca C, Denton CP, Gregersen PK, Agarwal S, Assassi S, Tan FK, Arnett FC, Radstake TR, Mayes MD, Martin J. Identification of novel genetic markers associated with clinical phenotypes of systemic sclerosis through a genome-wide association strategy. *PLoS Genet*. 2011;7(7): e1002178.
250. **Arnett FC**, Gourh P, Shete S, Ahn CW, Honey RE, Agarwal SK, Tan FK, McNearney T, Fischbach M, Fritzler MJ, Mayes MD, Reveille JD. Major histocompatibility complex (MHC) class II alleles, haplotypes and epitopes which confer susceptibility or protection in systemic sclerosis: analyses in 1300 Caucasian, African-American and Hispanic cases and 1000 controls. *Ann Rheum Dis*. 2010;69(5):822-7.
251. **Zhou X**, Lee JE, Arnett FC, Xiong M, Park MY, Yoo YK, Shin ES, Reveille JD, Mayes MD, Kim JH, Song R, Choi JY, Park JA, Lee YJ, Lee EY, Song YW, Lee EB. HLA-DPB1 and DPB2 are genetic loci for

systemic sclerosis: a genome-wide association study in Koreans with replication in North Americans. *Arthritis Rheum* 2009;60(12):3807–14.

252. **Radstake TR**, Gorlova O, Rueda B, Martin JE, Alizadeh BZ, Palomino-Morales R, Coenen MJ, Vonk MC, Voskuyl AE, Schuerwegh AJ, Broen JC, van Riel PL, van 't Slot R, Italiaander A, Ophoff RA, Riemekasten G, Hunzelmann N, Simeon CP, Ortego-Centeno N, González-Gay MA, González-Escribano MF; Spanish Scleroderma Group, Airo P, van Laar J, Herrick A, Worthington J, Hesselstrand R, Smith V, de Keyser F, Houssiau F, Chee MM, Madhok R, Shiels P, Westhovens R, Kreuter A, Kiener H, de Baere E, Witte T, Padykov L, Klareskog L, Beretta L, Scorza R, Lie BA, Hoffmann-Vold AM, Carreira P, Varga J, Hinchcliff M, Gregersen PK, Lee AT, Ying J, Han Y, Weng SF, Amos CI, Wigley FM, Hummers L, Nelson JL, Agarwal SK, Assassi S, Gourh P, Tan FK, Koeleman BP, Arnett FC, Martin J, Mayes MD. Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. *Nat Genet* 2010;42(5):426–9.
253. **Allanore Y**, Saad M, Dieudé P, Avouac J, Distler JH, Amouyel P, Matucci-Cerinic M, Riemekasten G, Airo P, Melchers I, Hachulla E, Cusi D, Wichmann HE, Wipff J, Lambert JC, Hunzelmann N, Tiev K, Caramaschi P, Diot E, Kowal-Bielecka O, Valentini G, Mouthon L, Czirják L, Damjanov N, Salvi E, Conti C, Müller M, Müller-Ladner U, Riccieri V, Ruiz B, Cracowski JL, Letenneur L, Dupuy AM, Meyer O, Kahan A, Munnich A, Boileau C, Martinez M. Genome-wide scan identifies TNIP1, PSORS1C1, and RHOB as novel risk loci for systemic sclerosis. *PLoS Genet* 2011;7(7), e1002091.
254. **Mayes MD**, Bossini-Castillo L, Gorlova O, Martin JE, Zhou X, Chen WV, Assassi S, Ying J, Tan FK, Arnett FC, Reveille JD, Guerra S, Teruel M, Carmona FD, Gregersen PK, Lee AT, López-Isac E, Ochoa E, Carreira P, Simeón CP, Castellví I, González-Gay MÁ; Spanish Scleroderma Group, Zhernakova A, Padyukov L, Alarcón-Riquelme M, Wijmenga C, Brown M, Beretta L, Riemekasten G, Witte T, Hunzelmann N, Kreuter A, Distler JH, Voskuyl AE, Schuerwegh AJ, Hesselstrand R, Nordin A, Airó P, Lunardi C, Shiels P, van Laar JM, Herrick A, Worthington J, Denton C, Wigley FM, Hummers LK, Varga J, Hinchcliff ME, Baron M, Hudson M, Pope JE, Furst DE, Khanna D, Phillips K, Schiopu E, Segal BM, Molitor JA, Silver RM, Steen VD, Simms RW, Lafyatis RA, Fessler BJ, Frech TM, Alkassab F, Docherty P, Kaminska E, Khalidi N, Jones HN, Markland J, Robinson D, Broen J, Radstake TR, Fonseca C, Koeleman BP, Martin J. Immunochip analysis identifies multiple susceptibility loci for systemic sclerosis. *Am J Hum Genet.* 2014;2;94(1):47-61.
255. **He D**, Wang J, Yi L, Guo X, Guo S, Guo G6, Tu W, Wu W, Yang L, Xiao R1, Li Y, Chu H, Lai S, Jin L, Zou H, Reveille JD, Assassi S, Mayes MD, Zhou X. Association of the HLA-DRB1 with scleroderma in Chinese population. *PLoS One.* 2014;3;9(9):e106939.

256. **Beretta L**, Rueda B, Marchini M, Santaniello A, Simeón CP, Fonollosa V, Caronni M, Rios-Fernandez R, Carreira P, Rodriguez-Rodriguez L; Spanish Systemic Sclerosis Group, Moreno A, López-Nevot MA, Escalera A, González-Escribano MF, Martin J, Scorza R. Analysis of Class II human leucocyte antigens in Italian and Spanish systemic sclerosis. *Rheumatology (Oxford)*. 2012;51(1):52-9.
257. **Assassi S**, Del Junco D, Sutter K, McNearney TA, Reveille JD, Karnavas A, et al. Clinical and genetic factors predictive of mortality in early systemic sclerosis. *Arthritis Rheum* 2009;61(10):1403–11.
258. **Simeon CP**, Fonollosa V, Tolosa C, Palou E, Selva A, Solans R, Armadans L, Moreno E, Marsal S, Vilardell M. Association of HLA class II genes with systemic sclerosis in Spanish patients. *J Rheumatol*. 2009; 36(12):2733-6.
259. **Zochling J**, Newell F, Charlesworth JC, Leo P, Stankovich J, Cortes A, Zhou Y, Stevens W, Sahhar J, Roddy J, Nash P, Tymms K, Rischmueller M, Lester S, Proudman S, Brown MA. An Immunochip-based interrogation of scleroderma susceptibility variants identifies a novel association at DNASE1L3. *Arthritis Res Ther*. 2014;16(5):438.
260. **Gilchrist FC**, Bunn C, Foley PJ, Lympay PA, Black CM, Welsh KI, du Bois RM. Class II HLA associations with autoantibodies in scleroderma: a highly significant role for HLA-DP. *Genes Immun*. 2001; 2(2):76-81.
261. **Tikly M**, Rands A, McHugh N, Wordsworth P, Welsh K. Human leukocyte antigen class II associations with systemic sclerosis in South Africans. *Tissue Antigens*. 2004; 63(5):487-90.
262. **Rodriguez-Reyna TS**, Mercado-Velázquez P, Yu N, Alosco S, Ohashi M, Lebedeva T, Cruz-Lagunas A, Núñez-Álvarez C, Cabiedes-Contreras J, Vargas-Alarcón G, Granados J, Zúñiga J, Yunis E. HLA Class I and II Blocks Are Associated to Susceptibility, Clinical Subtypes and Autoantibodies in Mexican Systemic Sclerosis (SSc) Patients. *PLoS One*. 2015 May 20;10(5): e 0126727.
263. **Gladman DD**, Kung TN, Siannis F, Pellett F, Farewell VT, Lee P. HLA markers for susceptibility and expression in scleroderma. *J Rheumatol*. 2005;32(8):1481-7.
264. **Takeuchi F**, Nakano K, Yamada H, Hong GH, Nabeta H, Yoshida A, Matsuta K, Bannai M, Tokunaga K, Ito K. Association of HLA-DR with progressive systemic sclerosis in Japanese. *J Rheumatol*. 1994;21(5):857-63.

265. **Kang SH**, Park MH, Song EY, Kang SJ, Lee EB, Song YW, Takeuchi F. Association of HLA class II genes with systemic sclerosis in Koreans. *J Rheumatol*. 2001; 28(7):1577-83.
266. **Louthrenoo W**, Kasitanon N, Wichainun R, Wangkaew S, Sukitawut W, Ohnogi Y, Nakaue N, Kuwata S, Takeuchi F. Association of HLA-DRB1*15:02 and DRB5*01:02 allele with the susceptibility to systemic sclerosis in Thai patients. *Rheumatol Int*. 2013; 33(8):2069-77.
267. **Azzouz DF**, Rak JM, Fajardy I, Allanore Y, Tiev KP, Farge-Bancel D, et al. Comparing HLA shared epitopes in French Caucasian patients with scleroderma. *PLoS One* 2012;7(5), e36870.
268. **Zhou X**, Lee JE, Arnett FC, Xiong M, Park MY, Yoo YK, Shin ES, Reveille JD, Mayes MD, Kim JH, Song R, Choi JY, Park JA, Lee YJ, Lee EY, Song YW, Lee EB. HLA-DPB1 and DPB2 are genetic loci for systemic sclerosis: a genome-wide association study in Koreans with replication in North Americans. *Arthritis Rheum*. 2009;60(12):3807-14.
269. **Wang J**, Guo X, Yi L, Guo G, Tu W, Wu W, Yang L, Xiao R, Li Y, Chu H, He D, Jin L, Mayes MD, Zou H, Zhou X. Association of HLA-DPB1 with scleroderma and its clinical features in Chinese population. *PLoS One*. 2014; 31;9(1): e87363.
270. **Kuwana M**, Inoko H, Kameda H, Nojima T, Sato S, Nakamura K, Ogasawara T, Hirakata M, Ohosone Y, Kaburaki J, Okano Y, Mimori T. Association of human leukocyte antigen class II genes with autoantibody profiles, but not with disease susceptibility in Japanese patients with systemic sclerosis. *Intern Med*. 1999;38(4):336-44.
271. **Kang SH**, Park MH, Song EY, Kang SJ, Lee EB, Song YW, Takeuchi F. Association of HLA class II genes with systemic sclerosis in Koreans. *J Rheumatol*. 2001; 28(7):1577-83.
272. **Zhou XD**, Yi L, Guo XJ, Chen E, Zou HJ, Jin L, Mayes MD, Assassi S, Wang JC. Association of HLA-DQB1*0501 with scleroderma and its clinical features in Chinese population. *Int J Immunopathol Pharmacol*. 2013;26(3):747-51.
273. **Kuwana M**, Okano Y, Kaburaki J, Inoko H. HLA class II genes associated with anticentromere antibody in Japanese patients with systemic sclerosis (scleroderma). *Ann Rheum Dis*. 1995; 54(12):983-7.

274. **Yunis EJ**, Larsen CE, Fernandez-Viña M, Awdeh ZL, Romero T, Hansen JA, Alper CA. Inheritable variable sizes of DNA stretches in the human MHC: conserved extended haplotypes and their fragments or blocks. *Tissue Antigens*. 2003; 62(1):1-20.
275. **Vigone B**, Santaniello A, Marchini M, Montanelli G, Caronni M, Severino A, Beretta L. Role of class II human leucocyte antigens in the progression from early to definite systemic sclerosis. *Rheumatology (Oxford)*. 2015;54(4):707-11.
276. **Takeuchi F**, Nakano K, Yamada H, Hong GH, Nabeta H, Yoshida A, Matsuta K, Bannai M, Tokunaga K, Ito K. Association of HLA-DR with progressive systemic sclerosis in Japanese. *J Rheumatol*. 1994;21:857–863.
277. **Arnett FC**, Howard RF, Tan F, Moulds JM, Bias WB, Durban E, Cameron HD, Paxton G, Hodge TJ, Weathers PE, Reveille JD. Increased prevalence of systemic sclerosis in a Native American tribe in Oklahoma. Association with an Amerindian HLA haplotype. *Arthritis Rheum*. 1996;39(8):1362-70).
278. **Tan FK**, Stivers DN, Arnett FC, Chakraborty R, Howard R, Reveille JD. HLA haplotypes and microsatellite polymorphisms in and around the major histocompatibility complex region in a Native American population with a high prevalence of scleroderma (systemic sclerosis). *Tissue Antigens*. 1999;53(1):74-80.
279. **Inamoto Y**, Martin PJ, Flowers ME, Lee SJ, Carpenter PA, Warren EH, Boeckh M, Storer BE, Levine D, Fan W, Zhao LP, Hansen JA. Genetic Risk Factors for Sclerotic Graft-Versus-Host Disease Blood; 2014;124(21).
280. **Steen VD**, Medsger TA Jr. Improvement in skin thickening in systemic sclerosis associated with improved survival. *Arthritis Rheum*. 2001;44(12):2828-35.
281. **Meyer JM**, Evans TI, Small RE, Redford TW, Han J, Singh R, Moxley G. HLA-DRB1 genotype influences risk for and severity of rheumatoid arthritis. *J Rheumatol*. 1999;26(5):1024-34.
282. **www.cancer.gov/about-cancer/treatment/clinical-trials**

283. **van den Hoogen F**, Khanna D, Fransen J, Johnson SR, Baron M, Tyndall A, Matucci-Cerinic M, Naden RP, Medsger TA Jr, Carreira PE, Riemekasten G, Clements PJ, Denton CP, Distler O, Allanore Y, Furst DE, Gabrielli A, Mayes MD, van Laar JM, Seibold JR, Czirjak L, Steen VD, Inanc M, Kowal-Bielecka O, Müller-Ladner U, Valentini G, Veale DJ, Vonk MC, Walker UA, Chung L, Collier DH, Ellen Csuka M, Fessler BJ, Guiducci S, Herrick A, Hsu VM, Jimenez S, Kahaleh B, Merkel PA, Sierakowski S, Silver RM, Simms RW, Varga J, Pope JE. 2013 classification criteria for systemic sclerosis: an American college of rheumatology/European league against rheumatism collaborative initiative. *Ann Rheum Dis*. 2013;72:1747-55.
284. **Clements P**, Lachenbruch P, Siebold J, White B, Weiner S, Martin R, Weinstein A, Weisman M, Mayes M, Collier D, et al. Inter and intraobserver variability of total skin thickness score (modified Rodnan TSS) in systemic sclerosis. *J Rheumatol*. 1995;22(7):1281-5.
285. **Rogov SI**, Nekrasov AN. A numerical measure of amino acid residues similarity based on the analysis of their surroundings in natural protein sequences. *Protein Eng*. 2001;14(7):459-63.
286. **Taylor WR**. The classification of amino acid conservation. *J Theor Biol*. 1986;119:205–221.
287. **Dai S**, Murphy GA, Crawford F, Mack DG, Falta MT, Marrack P, Kappler JW, Fontenot AP. Crystal structure of HLA-DP2 and implications for chronic beryllium disease. *Proc. Natl. Acad. Sci. USA*. 2010;107:7425–7430
288. **Boin F**, Rosen A. Autoimmunity in systemic sclerosis: current concepts. *Curr Rheumatol Rep*. 2007; 9(2):165-72.
289. **Fava A**, Cimbri R, Wigley FM, Liu QR, Rosen A, Boin F. Frequency of circulating topoisomerase-I-specific CD4 T cells predicts presence and progression of interstitial lung disease in scleroderma. *Arthritis Res Ther*. 2016;4;18(1):99.
290. **Hu PQ**, Oppenheim JJ, Medsger TA Jr, Wright TM. T cell lines from systemic sclerosis patients and healthy controls recognize multiple epitopes on DNA topoisomerase I. *J Autoimmun*. 2006;26(4):258-67.

291. **Nihtyanova SI**, Denton CP. Autoantibodies as predictive tools in systemic sclerosis. *Nat Rev Rheumatol*. 2010; 6(2):112-6.
292. **Baker KS**, Loberiza FR Jr, Yu H, Cairo MS, Bolwell BJ, Bujan-Boza WA, Camitta BM, Garcia JJ, Ho WG, Liesveld JL, Maharaj D, Marks DI, Schultz KR, Wiernik P, Zander AR, Horowitz MM, Keating A, Weisdorf DJ. Outcome of ethnic minorities with acute or chronic leukemia treated with hematopoietic stem-cell transplantation in the United States. *J Clin Oncol*. 2005;1;23(28):7032-42.