Online Supplemental Methods.

Human subjects and skin samples. Patients with recent morphea lesions of cGVHD (n=5) and patients with lichen planus lesions of cGVHD (n=8), all under minimal medical treatment, diagnosed according to the standard National Institutes of Health consensus criteria with histologic confirmation, were recruited from the Hematology Transplantation Department of Saint-Louis Hospital (Paris, France), and were enrolled in the study between January 2017 and February 2019. Supplemental table S1 summarizes the clinical characteristics of these patients. Two skin biopsies were performed in affected area using a 3 to 5-millimeter skin punch biopsies under local anesthesia from each patient. A skin sample was place in formalin and embedded in paraffin for histopathologic evaluation. The other skin sample underwent excision of the hypoderm and the dermoepidermal part was stored in 400µl of RNAlater Stabilization Solution at −80 °C for RNA sequencing.

Six healthy controls skin samples were obtained from healthy individuals undergoing plastic surgery using a 3 to 5-millimeter skin punch biopsies. Hypoderm was removed, and the dermoepidermal portion was stored in 400µl of RNAlater Stabilization Solution at −80 °C.

RNA extraction. Total RNA was isolated from skin samples using the RNeasy Plus Universal Kit (Qiagen, Hilden, Germany) according to the manufacturer's specifications[1]. Preceding RNA isolation all samples were disrupted using safe lock tubes with a 5 mm stainless steel beads (Qiagen) in a TissueLyser II (25Hz for 10 minutes) together with QIAzol Lysis Reagent (Qiagen). gDNA eliminator solution was used to remove genomic DNA contamination. Next, 200µl of chloroform was added to each tube, mixed by shaking up and down and centrifuged at 4°C. The upper aqueous layer containing DNA and RNA was carefully collected and several purification steps were performed using ethanol 96% and 80%, RWT and RPE buffer (Qiagen), columns (Qiagen) and centrifugation. Elution of RNA was performed with RNase-free water. The extracted RNA was then stored at -80°C.

Sample quality assessment. Initial quality control was performed at the National Institute of Health and Medical Research (INSERM, Paris, France). RNA isolates were quantified using spectrophotometry with NanoDrop 2000 (Thermo Scientific). The degree of sample contamination by nucleic acids, proteins and other contaminants was evaluated by determining the ratios of sample absorbance at 260 and 280 nm (A260/280). Values ≥1.8 were considered "pure" and acceptable for downstream molecular analyses.

In addition, complete RNA quality control on each sample was performed at CNRGH (Centre National de Recherche en Génomique Humaine, Institut de Biologie François Jacob, Evry, FRANCE). RNA concentrations were measured by UV quantification on a NanoDrop™ 8000 spectrophotometer (in duplicate). Indeed, RNA quality was tested using RNA Integrity Number (RIN) measurements by running a sample aliquot on a Bioanalyzer 2100 from Agilent, using the RNA6000 Nano Labchip kit (#5065-4476, Agilent Technologies, Inc., Santa Clara, CA)[2]. Selected samples for RNA sequencing were of good quality (no DNA contamination, no important degradation, RIN value > 7) and had concentration between 30-150 ng/µL.

RNA library preparation and sequencing

RNA sequencing was performed at CNRGH. Libraries have been prepared using the "TruSeq stranded mRNA" Kit from Illumina, which selects polyA+ RNAs as a first step of library preparation and allows to get a clear view of the protein-coding transcriptome with strand-specific information. An input of 1 µg total RNA was used for all samples, and libraries were prepared on an automated platform, according to manufacturer's instructions. Library quality and concentration have been checked by LabGx (Perkin Elmer) and sample libraries have then been pooled before sequencing to reach the expected sequencing depth. Sequencing has been performed on an Illumina HiSeq4000 as paired-end 101 bp reads, using Illumina sequencing reagents. Libraries were pooled by 6 samples per lane, corresponding on average to 40 to 50 million sequenced fragments (or 80 to 100 million total reads).

The raw sequencing data was stored in FastQ format.

Quality control of RNAseq data was performed at CNRGH. Fastq files have been be processed by in-house CNRGH tools in order to assess quality of raw and genomic-aligned nucleotides. The following steps have been performed on a random selection of 2 x 10 million reads including the following steps : *(i)* a cleanup of FASTQ files using Trimmomatic tool in order to determine the percentage of reads remaining after the removing of adapters (sequencing and multiplexing) and lower quality sequences; *(ii)* an alignment on the reference genome (Hisat2) as well as on the transcriptome and the ribosomal RNA (Bowtie2); *(iii)* the use of RSeqC and picardTools tools in order to generate quality metrics such as the percentage of "mapping" on the genome and transcriptome, the percentage of duplicate sequences, the percentage of ribosomal RNA (rRNA) and the total number of sequences.

RNA-seq data processing for expression analysis

Gene expression analysis was performed at Fondation Jean Dausset-CEPH (Centre d'Étude du Polymorphisme Humain, CEPH, Paris, France). Cleaned reads (FASTQ files) obtained for each sample were aligned to the human reference genome GRCh38_r79 using STAR (version 2.5.3a)[3,4]
("*Genome Reference Consortium, human version 38*" (ftp://ftp.ensembl.org/pub/release-79/fasta/homo_sapiens/dna/Homo_sapiens.GRCh38.dna.chromosome.{1..22}.fa.gz/ ftp://ftp.ensembl.org/pub/release-79/fasta/homo_sapiens/dna/Homo_sapiens.GRCh38.dna.chromosome.{MT,X,Y}.fa.gz) and gene annotation file GRCh38.79.gtf (ftp://ftp.ensembl.org/pub/release-79/gtf/homo_sapiens/Homo_sapiens.GRCh38.79.gtf.gz). We used for downstream analyzes the gene counts generated by STAR.

The differential expression analysis was carried out at gene and transcript levels , between 2 conditions (normal skin *versus* damaged skin (Morphea/Lichen)) using *DESeq2* package (version: 1.26.0)[5] implemented in R (version R 3.4 available on the Bioconductor website: https://support.bioconductor.org) with its graphical user interface Rstudio (Version 1.2.2019). *DESeq2* enables the identification of differentially expressed genes (DEGs) whose expression abundance is significantly increased (up-regulated) or decreased (down-regulated) in a pathological condition compared to a control condition. The results of the differential expression were defined for each gene in fold change (FC), p-value (estimated by Wald test) and adjusted p-value (padj) computed by Benjamini-Hochberg[6] calculation to correct for multiple testing. The obtained DEGs were identified according to the ENSEMBL nomenclature and matches in "hgnc_symbol" were also used.

Pathway enrichment analyzes of DEGs

The functional analyzes of the DEGs making it possible to identify enriched biological pathways between the two patient groups (lichenous and morphea), were performed at CEPH using "Gene Set Enrichment Analysis" (GSEA)[7] and "Ingenuity Pathway Analysis" (IPA) software[8].

1- Gene Set Enrichment Analysis (GSEA)

The GSEA developed at the Broad institute (http://www.broad.mit.edu/gsea/index.jsp) allows the functional characterization of differentially expressed transcripts in one condition *versus*

another condition using biological pathway annotation. Data from normalized read counts per gene obtained by *DESeq2* were used as input to GSEA. GSEA calculates an enrichment score (ES) using a weighted Kolmogorov–Smirnov-like statistic. The statistical significance (nominal P value) of the ES was estimated by using an empirical phenotype-based permutation test procedure that preserves the complex correlation structure of the gene expression data. It first normalizes the ES for each gene set to account for the size of the set, yielding a normalized enrichment score (NES). NES reflects the degree of overrepresentation of a set of genes in a list of ranked genes. GSEA also controls the proportion of false positives by calculating the false discovery rate (FDR) corresponding to each NES; it is computed by comparing the tails of the observed and null distributions for the NES (1000 permutations). The latest gene sets collection of the "Molecular Signatures Database" v7.0. (MsigDB) (http://www.broadinstitute.org/gsea/msigdb) was used for the analysis. The selection criteria used for "pathways" are as follows: nominal p-values and the lowest False Discovery Rate (FDR) (< 5%) and values of NES ≥ 1.5. The threshold of significance of NES (≥ 1.5) has been defined arbitrarily. An enrichment graph was generated for each enriched channel, making it possible to visualize the enrichment score, namely the genes the most involved in the enrichment of the channel in question. GSEA also makes it possible to draw up a "heatmap" highlighting the 100 most differentially expressed genes on the basis of the counting data of reads standardized by *DESEq2*.

2- Ingenuity Pathway Analysis (IPA)

We used the commercial QIAGEN's *Ingenuity® Pathway Analysis* (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity) software for functional pathway and upstream regulatory analysis (URA) of DEGs identified in this study. IPA is an enrichment software for ontological functions and signaling pathways and was used to define function and classify genes as well as the biological significance of differentially expressed genes (DEGs) in each subset of patients (Morphea/Lichen). *DEGs* identified by *DESEQ2* were imported to IPA. In the canonical pathway analysis, IPA calculates a p value for each pathway, indicating that the probability of association between the DEGs and the canonical path is not random. A threshold value of p-value was set at less than 0.05, namely -log10 (p-value)> 1.3. The p-values were calculated by Fisher's exact test and corrected for multiple testing using Benjamini-Hochberg (B-H) procedure to define the meaning of the channels associated with our DEGs data. In addition, DEGs were also analyzed for the prediction of activated or inhibited canonical pathways on the basis of z-score. IPA automatically calculates the z-score based on the DEGs.

A positive z-score suggests prediction of activation, while a negative z-score indicates prediction of inactivation of the pathway. IPA calculates B-H corrected p and z-score values from the DEGs data.

Immunohistochemistry (IHC) staining

The specimens were fixed in formalin and embedded in paraffin (FFPE). After antigen retrieval in citrate buffer and heating in a microwave oven, 5-µm thick sections were incubated with anti-humanTREM-1 polyclonal antibody (AF1278, R&D Systems) raised in goat at 1:50 dilution. The sections were revealed by using a 3 step-technique with ABC complexes and peroxidase as previously described9.

Real-Time quantitative PCR

Real-Time quantitative PCR was carried out using the HotStarTaq DNA polymerase kit (Qiagen) using a LightCycler 480 II thermocycler (Roche) after reverse transcription from 150 ng total RNA (PrimeScript RT Reagent Kit with gDNA Eraser (Perfect Real Time) Takara). TREM1 primer was obtained from Qiagen. GAPDH was used as reference gene for normalization and primer was purchased from TIB MOLBIOL (Berlin, Germany) and was as followed(listed5’→3’):TGCACCACCAACTGCTTAGC(forward), GGCATGGACTGTGGTCATGAG (reverse).

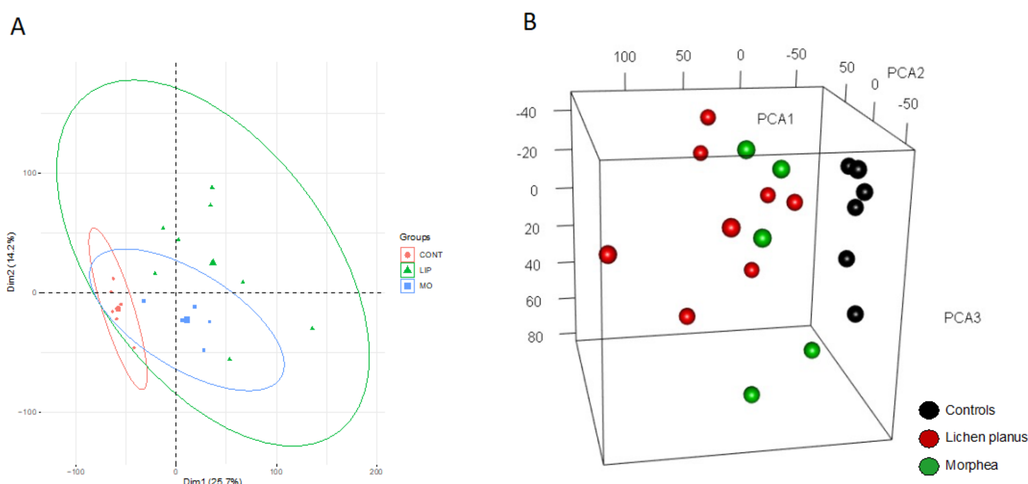
PCR condition were 95° for 10 minutes followed by 50 cycles of 95°C for 30 seconds, 58°C for 30 seconds and 72° C for 20 seconds. At the end of the amplification reaction, melting curve examination was performed to confirm the specificity as well as the integrity of the PCR product by the presence of a single peak. The relative expression levels of mRNA were determined using the dCT formula and fold changes were calculated as 2-dCT. Inferences between samples were made using non-parametric test (Wilcoxon rank-sum test, Kruskal-Wallis H-test).

Study approval. This study was reviewed and approved by the local ethics committee at CPP Paris Ile de France IV (Paris, France). All samples were obtained from patients recruited from the Hematology Transplantation Department of Saint-Louis Hospital in Paris, France and were enrolled in this study between January 2017 and February 2019. Written, informed consent was obtained from these patients and recorded in their medical file. The study was conducted according to the principles of the Declaration of Helsinki.

References

1. Sellin Jeffries MK, Kiss AJ, Smith AW, Oris JT. A comparison of commercially-available automated and manual extraction kits for the isolation of total RNA from small tissue samples. *BMC Biotechnol.* 2014;14:94.
2. Davies J, Denyer T, Hadfield J. Bioanalyzer chips can be used interchangeably for many analyses of DNA or RNA. *BioTechniques*. 2016;60(4):197–199.
3. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21.
4. Dobin A, Gingeras TR. Mapping RNA-seq Reads with STAR. *Curr Protoc Bioinformatics*. 2015;51:11.14.1-11.14.19.
5. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
6. Benjamini Y, Hochberg Y. Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing. *J. Royal Statist. Soc., Series B*. 1995;57:289–300.
7. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* 2005;102(43):15545–15550.
8. Krämer A, Green J, Pollard J, Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics*. 2014;30(4):523–530.
9. Boufenzer A, Lemarié J, Simon T, et al. TREM-1 Mediates Inflammatory Injury and Cardiac Remodeling Following Myocardial Infarction. *Circ Res*. 2015;116(11):1772–1782.

Supplemental Figure S1. Two (A) and three (B) principal component analysis (PCA) of the whole transcriptome showing clear separation of healthy controls from the lesional skin samples however, without perfect separation between the two subtypes of cutaneous cGVHD.



Supplemental Table S1. Sample description. Clinical and molecular features collected for each patient/sample used in the RNA-seq analysis.

	Gender	Age	Biopsy	Time from HSCT to sample (days)	Type	GVHD Score skin (NIH)	Current therapy
LP1	F	26	Upper leg	554	LP	Moderate	Systemic corticotherapy (0.08 mg/kg/j)
LP2	M	38	Back	54	LP	Severe	Tacrolimus (0.06 mg/kg/j) Systemic corticotherapy (0.7mg/kg/j) Thymoglobulin 1 month prior
LP4	F	66	Axillae	2073	LP	Moderate	None
LP5	F	42	Trunk	162	LP	Severe	None
LP6	F	57	Trunk	186	LP	Severe	None
LP7	F	60	Lower leg	64	LP	Moderate	Substitutive Immunoglobulin
LP8	M	67	ND	ND	LP	Moderate	None
LP9	F	59	ND	ND	LP	Moderate	None
Mo1	M	59	Trunk	4546	Morphea	Severe	Systemic corticotherapy (0.05mg/kg)
Mo2	M	55	Lower arm	712	Morphea	Severe	None
Mo3	F	42	Upper leg	1121	Morphea	Severe	Topical corticotherapy
Mo4	M	55	ND	1288	Morphea	Moderate	None
Mo5	M	67	ND	ND	Morphea	Severe	Topical corticotherapy
CONT1	M	31	Abdomen	N/A	HC	N/A	None
CONT2	F	18	Breast	N/A	HC	N/A	None
CONT3	F	47	Breast	N/A	HC	N/A	None
CONT4	F	56	Abdomen	N/A	HC	N/A	None
CONT5	F	23	Breast	N/A	HC	N/A	None
CONT6	F	19	Breast	N/A	HC	N/A	None

ND, no data; N/A, not applicable

Supplemental Table S2 (Excel). Differentially expressed genes (DEGs). (A) Lichen planus cGVHD versus control. (B) Morphea cGVHD versus control. (C) Lichen planus versus morphea. (D) Common DEGs shared by lichen planus and morphea cGVHD.

Supplemental Table S3. TREM-1 and GAPDH mRNA expression in lichen planus cGVHD, morphea cGVHD and healthy controls.

	GAPDH (measure #1)	GAPDH (measure #2)	GAPDH (mean)	TREM-1 (measure #1)	TREM-1 (measure #2)	TREM-1 (mean)	TREM-1 (2^{-DCT})	Relative mRNA expression normalized to GAPDH
LP1	20.69	21.53	21.11	31.03	30.55	30.79	0.001219073	0.00121907
LP2	21.11	20.98	21.045	31.53	31.79	31.66	0.000637627	0.00063763
LP4	21.87	21.15	21.51	31.52	31.86	31.69	0.000862015	0.00086201
LP5	21.23	20.94	21.085	33.11	32.14	32.625	0.000335826	0.00033583
LP6	22.4	23.15	22.775	35.96	35.63	35.795	0.00012039	0.00012039
LP7	22.6	21.36	21.79	34.97	34.73	34.85	0.000117098	0.0001171
Mo1	21.98	22.02	22	35.86	35.75	35.805	6.98684E-05	6.9868E-05
Mo2	22.71	22.37	22.54	34.05	33.6	33.825	0.000400753	0.00040075
Mo3	22.07	21.81	21.94	36.57	36.77	36.67	3.67983E-05	3.6798E-05
Mo4	22.08	21.91	21.995	34.77	35.43	35.1	0.000113502	0.0001135
Mo5	23.53	23.28	23.405	34.27	34.26	34.265	0.00053804	0.00053804
CONT1	22.02	21.92	21.97	37.92	38.92	38.105	1.38957E-05	1.3896E-05
CONT2	22.99	22.91	22.95	36.36	35.43	35.895	0.000126814	0.00012681
CONT3	22.64	22.89	22.765	36.56	38.17	37.365	4.0268E-05	4.0268E-05
CONT4	22.83	22.39	22.61	35.04	34.63	34.835	0.000208886	0.00020889
CONT5	23.22	23.33	23.275	45	38.18	41.59	3.06645E-06	3.0664E-06