Appendix S1

SUPPLEMENTARY MATERIALS AND METHODS

Patient characteristics

The patients enrolled as a part of this study were unrelated Caucasians living in Estonia. The patient volunteers were recruited from the outpatient clinic as they arrived for appointments, with no discrimination with regards to the clinical characteristics of the disease, in order to minimize the effect of favouring patients with clear-cut phenotypes. The mean ± standard deviation (SD) of the age of patients were 44 ± 12 and 45 ± 19 years for the psoriasis vulgaris (PsV) and atopic dermatitis (AD) group, respectively. The male-to-female ratio was 2.3 in the PsV group and 0.5 in the AD group. The overall male-to-female ratio was 1.1. Disease severity was quantified as Psoriasis Area and Severity Index (PASI) score for PsV patients and as Eczema Area and Severity Index (EASI) score for AD patients. Detailed characteristics of recruited patients are presented in **STable I**. Characteristics of the patients included in the GSE66511 RNA-seq dataset and qPCR dataset 1 can be found in the study published by Keermann et al. (21).

Quantitative PCR methods

Fastprep-24 instrument with lysing matrix D tubes (MP Biomedicals, Santa Ana, CA, USA) was used to homogenize biopsy specimens. Total RNA was extracted and DNase treated with RNeasy Fibrous Tissue Mini Kit (Qiagen, Valencia, CA, USA) according to the manufacturer's protocol. Isolated RNA was dissolved in RNase-free water and stored at –80°C. The quality of total RNA was evaluated with Agilent 2100 Bioanalyzer and the RNA 6000 Nano Kit (Agilent Technologies Inc., CA, USA). The RNA integrity number (RIN) of all samples was ≥7. The quantity was measured applying Qubit 2.0 fluorometer and Qubit RNA HS Assay kit (Thermo Fisher Scientific Inc., CA, USA).

Primers were designed using Primer3 software v4.1.0 (S1). Detailed information for the primers used in the qPCR experiments is presented in **STable II**. *In silico* primer specificity was screened with GenomeTester 1.3 (S2). The specificity was also confirmed by melting curve analysis and Sanger sequencing.

The qPCR experiments were conducted according to the MIQE guidelines (22). 250 ng of total RNA was used with High Capacity RNA-to-cDNA Kit (Life Technologies Co, USA) for cDNA

STable I. Characteristics of the recruited patients (qPCR dataset 2)

Sample Code	Disease	Sex	Age, years	Disease duration	Disease severity	Nail involvement	PsA
1	PsV	F	57	20	20	✓	✓
2	PsV	M	39	26	26	✓	
3	PsV	M	31	5	20	✓	
4	PsV	F	33	17	24	✓	
5	PsV	M	71	40	18		
6	PsV	F	45	31	13		
7	PsV	M	41	14	21	✓	✓
8	PsV	M	46	25	20	✓	✓
9	PsV	M	32	20	23	✓	✓
10	PsV	F	66	50	21		
11	PsV	M	53	32	18		
12	PsV	M	46	3	25		
13	PsV	M	26	4	27		
14	PsV	M	50	20	21	✓	
15	PsV	M	39	13	*NA**	✓	
16	PsV	F	35	17	20		
17	PsV	M	34	26	21		
18	PsV	M	35	5	16		
19	PsV	F	57	2	12		
20	PsV	M	47	20	21		
						AD in family	Asthma
21	AD	F	35	35	23	✓	
22	AD	F	41	41	28	✓	
23	AD	M	74	1	25		
24	AD	F	55	54	26	✓	
25	AD	F	47	2	21		
26	AD	M	27	1	22		
27	AD	F	65	1	21		
30	AD	F	60	3	22		
31	AD	F	40	10	8		
32	AD	F	40	20	21		
33	AD	M	21	1	22	✓	✓
28	AD	M	22	2	24	✓	
29	AD	M	57	1	32		
34	AD	F	80	40	10		
36	AD	F	26	NA	15	✓	
37	AD	F	20	NA	22	✓	✓
38	AD	M	64	NA	8		
35	AD**	F	34	8	9		

*Missing data; **Dry subtype of atopic dermatitis.
Disease severity was quantified as Psoriasis Area and Severity Index (PASI) score for Ps patients and as Eczema Area and Severity Index (EASI) score for patients with AD. Disease duration is marked as "NA" if recruitment took place during first diagnosis of the disease. AD: atopic dermatitis; PsV: psoriasis vulgaris; PsA: psoriatic arthritis; M: male; F: female.

STable II. Quantitative PCR validation

Gene symbol	Sequence Acession No	PCR forward (upper) and reverse (lower) primer sequences (5'–3')	Amplicon length (bp)	Location primer	Targeted splice variants (Transcript ID)
IGFL1	ENSG00000188293	GGCTGCATCGTAGCTGTCTT	167	Exon 1–2	ENST00000437936.1
		ACGGCATCATCATAGCAACA		Exon 3	
CCL27	ENSG00000213927	CAGCTCTACCGAAAGCCACT	173	Exon 2	ENST00000259631.4
		TCTTGGTGCTCAAACCACTG		Exon 3	
NOS2	ENSG00000007171	CCATAAGGCCAAAGGGATTT	109	Exon 4–5	ENST00000313735.10
		ATCTGGAGGGGTAGGCTTGT		Exon 5	ENST00000621962.1
C10orf99	ENSG00000188373	TCCTGCTTCTCTGCTTCTCC	145	Exon 1	ENST00000372126.3
		CAGAGCCTCACATGATGTCC		Exon 3	
IL36G	ENSG00000136688	TTTGGGAATCCAGAATCCAG	140	Exon 4	ENST00000259205.4
		TGGCACGGTAGAAAAGGAAG		Exon 5	ENST00000376489.6
HPRT1	ENSG00000165704	GACTTTGCTTTCCTTGGTCAGG	101	Exon 6	ENST00000298556.7
		AGTCTGGCTTATATCCAACACTTCG		Exon 7	ENST00000462974.5
					ENST00000475720.1
ALAS1	ENSG00000023330	CAGCCACATCATCCCTGTG	115	Exon 10	ENST00000394965.6
		GGGCACCGTAGGGTAATTG		Exon 11	ENST00000484952.5
					ENST00000310271.6
					ENST00000469224.5
					ENST00000493402.1
					ENST00000459884.1
OSBP	ENSG00000110048	GGAGCATCATGAAGAACTGC	197	Exon 7	ENST00000263847.5
		GGAGTAGGAGGACACGGTGA		Exon 8	ENST00000525357.1
					ENST00000528903.1

synthesis. QuantiTect Reverse Transcription Kit (Qiagen) was applied for cDNA synthesis for calibration sample (CS) (pooled RNA from psoriasis lesional skin and control skin). Both cDNA synthesis kits were used according to the manufacturer's protocol. The PCR inhibition was tested with different cDNA dilutions and 5 times dilution (final amount in PCR reaction is 2.5 ng) was the optimal. cDNA was used as a template for qPCR in Quantstudio 12k Flex Real-Time PCR system platform (Thermo Fisher Scientific Inc., CA, USA).

The qPCR conditions for all 5 target genes and 3 reference genes (for all genes the amplification efficiencies were 100 ± 20%) were the same. The qPCR was conducted in 4 replicates and with reaction volume 10 µl on 384 plate format using EvaGreen qPCR Supermix (Solis BioDyne, Tartu, Estonia), final primer concentration 400 nM and cDNA input 2.5 ng (5× dilution) per reaction. The qPCR programme was as follows: hold stage – 95°C, 15 min; 40 cycles of PCR stage – 95°C, 20 s; 60°C, 20 s; 72°C, 20 s, melt curve stage –95°C, 20 s; 60°C, 20 s; 95°C, 20 s. The information about the qPCR validation are shown in **STable III** and **SFig. 1**.

Relative expression levels of the marker genes were obtained by normalizing the quantitation cycle (Cq) values against the geometrical average of 3 reference genes (*HPRT1, ALAS1, OSBP*) and the calibration sample in order to account for intra- and interplate variation, resulting in $-\Delta\Delta$Cq values. All the measurements were performed in 4 technical replicates and the mean of all replicates was used in the aforementioned calculations. As the expression level of *IL36G, NOS2, IGFL1* and *C10orf99* in psoriasis nonlesional skin and control samples, and expression of *CCL27* in psoriasis lesional skin samples was very low (Cq >30), often it was not possible to get results for all 4 replicates. Thus, in some cases only 1 replicate was considered as valid, and the SD was also high (up to 2.47). In other cases (Cq <30), SD <0.16 was considered as acceptable.

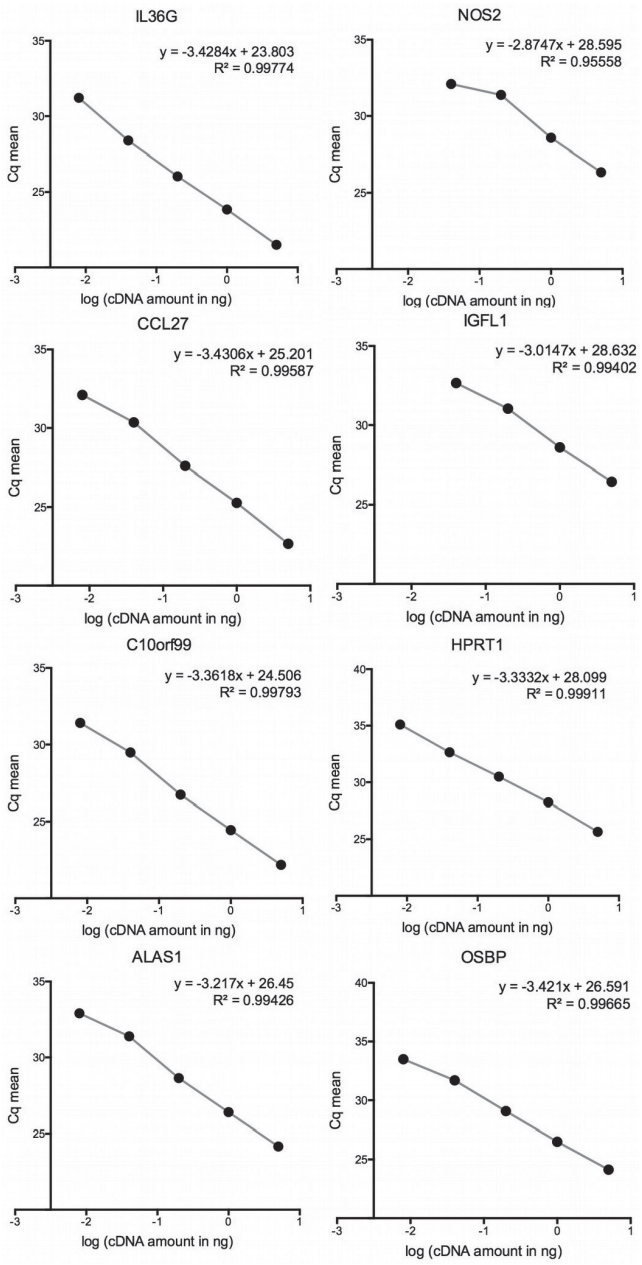
Cq values were taken as the average of the 4 (or less) technical replicates. The formulas for the expression analysis were as follows:

$C_{qreferencegenes}$ = Geometrical average of Cq from 3 reference genes

$\Delta Cq_{targetgene} = Cq_{targetgene} - Cq_{referencegenes}$ for patients' and controls' samples and calibration sample

$\Delta\Delta Cq_{targetgene} = \Delta Cq_{targetgene\ (patients'/controls'\ samples)} - \Delta Cq_{targetgene\ (calibration\ sample)}$

$-\Delta\Delta Cq_{targetgene}$ values were used for downstream analysis and visualization.



SFig. 1. Calibration curves for target and reference genes. Both the formula of th calibration curve and the correlation coefficient are shown.

STable III. Quantitative PCR validation

Gene symbol	Specificity confirmed by GenomeTester 1.3	Cq (SD) of negative control	Sample type used for calibration curve*	Formula for calibration curve**	PCR efficiency calculated from slope (%)	r^2 of calibration curve	Linear dynamic range (cDNA amount in ng)	Cq variantion at LOD as SD	$LOD_{relative\ (ng)}$
IGFL1	NO	NA	PP1	y=–3.0147x + 28.632	114.7	0.99402	0.04 – 5.0 ng	0.57	0.04
CCL27	Additional product from chr 9 (377 bp)	37.50 (1.69)	CP1	y=–3.4306x + 25.201	95.7	0.99587	0.008 – 5.0 ng	0.29	0.008
NOS2	NO	36.22 (1.43)	PP1	y=–2.9335x + 28.969	119.3	0.99837	0.04 – 5.0 ng	0.3	0.04
C10orf99	NO	37.41 (0.82)	PP1	y=–3.3618x + 24.506	98.4	0.99793	0.008 – 5.0 ng	0.07	0.008
IL36G	NO	38.63 (NA)	PP1	y=–3.4284x + 23.803	95.8	0.99774	0.008 – 5.0 ng	0.1	0.008
HPRT1	NO	35.92 (NA)	PP1	y=–3.3332x + 28.099	99.5	0.99911	0.008 – 5.0 ng	0.51	0.008
ALAS1	Additional product from chr 3 (821 bp)	35.49 (NA)	PP1	y=–3.217x + 26.45	104.4	0.99426	0.008 – 5.0 ng	0.44	0.008
OSBP	NO	35.05 (0.52)	PP1	y=–3.421x + 26.591	96.0	0.99665	0.008 – 5.0 ng	0.42	0.008

NO: no additional PCR product from genomic DNA according to GenomeTester software; Cq: crossing point; SD: standard deviation; PP1: pool of psoriasis lesional skin RNA; CP1: pool of control skin RNA; r2: correlation coefficient; LOD: limit of detection.

STAR parameters

--genomeDir GRCh38_Gencode26
--outFilterType BySJout
--outFilterMultimapNmax 20
--alignSJoverhangMin 8
--alignSJDBoverhangMin 1
--outFilterMismatchNmax 999
--outFilterMismatchNoverReadLmax 0.04
--alignIntronMin 20
--alignIntronMax 1000000
--alignMatesGapMax 1000000
--quantMode GeneCounts

Transformation of gene expression values

As qPCR, RNA-seq, and microarray data were used interchangeably to test the final support vector machine (SVM) model, gene expression values were transformed to new features in order to overcome inter-dataset differences caused by both batch effects and the use of different expression quantification technologies. The marker genes were ordered on the x-axis and coded as integers, as follows: *IL36G* as 1, *CCL27* as 2, *NOS2* as 3, and *C10orf99* as 4. In case of other sets of marker genes, the orders of genes were as follows: (a) *CCL27* as 1 and *NOS2* as 2; (b) *IL36G* as 1, *CRABP2* as 2, *S100A7A* as 3, and *IL36RN* as 4; (c) *SPRR2A* as 1, *PRELP* as 2, *ARG1* as 3, and *KYNU* as 4. The expression values of corresponding genes formed the y-axis as RPKM, –ΔΔCq or Cy5/Cy3 values. Next, a linear model was fitted to connect the expression values along the x-axis with a degree 3 polynomial line. This was done by using the lm() function in R. The resulting polynomial line equation intercept (β0) and coefficients ($\beta 1$, $\beta 2$, and $\beta 3$) as in $y=\beta 0 + \beta 1\ x + \beta 2\ x^2 + \beta 3\ x^3 + \varepsilon 1$ formed the 4 new features. The datasets were grouped by gene expression quantification method and the new features were then standardized to z-scores, where standard deviation (SD) and mean were calculated separately for each gene and quantification method (RNA-seq, microarray or qPCR). The GSE66511 RNA-seq dataset containing measurements for the same samples as the qPCR dataset 1 was included into the test set as well in order to control for potential errors resulting from feature transformation. The aforementioned data transformation minimizes the inter-dataset differences resulting from the use of different expression quantification methods and batch effects. As an alternative strategy, the polynomial transformation was omitted and the gene and method-wise z-scores were calculated directly from gene expression values (RPKM, –ΔΔCq or Cy5/Cy3).

SUPPLEMENTARY REFERENCES

S1. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3 – new capabilities and interfaces. Nucleic Acids Res 2012; 40: e115.
S2. Andreson R, Reppo E, Kaplinski L, Remm M. GENOMEMASKER package for designing unique genomic PCR primers. BMC Bioinformatics 2006; 7: 172.