Genomics Data 4 (2015) 115–118

Contents lists available at ScienceDirect

Genomics Data

journal homepage: http://www.journals.elsevier.com/genomics-data/

Data in Brief

A microarray analysis of two distinct lymphatic endothelial cell populations

Bernhard Schweighofer [a], Sabrina Rohringer [b,c], Johannes Pröll [d], Wolfgang Holnthoner [b,c,*]

[a] Skin and Endothelium Research Division, Department of Dermatology, Medical University of Vienna, Vienna, Austria
[b] Ludwig Boltzmann Institute for Experimental and Clinical Traumatology, Vienna, Austria
[c] Austrian Cluster for Tissue Regeneration, Vienna, Austria
[d] Red Cross Blood Transfusion Service of Upper Austria, Linz, Austria

ARTICLE INFO

Article history:
Received 15 March 2015
Received in revised form 2 April 2015
Accepted 3 April 2015
Available online 11 April 2015

ABSTRACT

We have recently identified lymphatic endothelial cells (LECs) to form two morphologically different populations, exhibiting significantly different surface protein expression levels of podoplanin, a major surface marker for this cell type. *In vitro* shockwave treatment (IVSWT) of LECs resulted in enrichment of the podoplanin^{high} cell population and was accompanied by markedly increased cell proliferation, as well as 2D and 3D migration. Gene expression profiles of these distinct populations were established using Affymetrix microarray analyses. Here we provide additional details about our dataset (NCBI GEO accession number GSE62510) and describe how we analyzed the data to identify differently expressed genes in these two LEC populations.

Specifications	
Organism/cell line/tissue	*Homo sapiens*/immortalized lymphatic endothelial cells isolated from foreskin
Sex	*Male*
Sequencer or array type	*3' IVT Expression Analysis on Affymetrix GeneChip Human Genome U133 Plus 2.0 array*
Data format	*Raw data: CEL files and RMA normalized*
Experimental factors	*Flow cytometry-sorted LEC population 1 versus population 2*
Experimental features	*Immortalized lymphatic endothelial cells from foreskin were flow-sorted according to differences in FSC and SSC values*
Consent	*Cells were isolated from healthy donors with authorization of a local ethics committee and informed consent by the donor*
Sample source location	*Vienna, Austria*

Direct link to deposited data

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE62510

Experimental design, materials and methods

Lymphatic endothelial cell isolation

Cells were isolated from healthy donors with authorization of a local ethics committee and informed consent by the donor. LECs were isolated from human foreskins *via* podoplanin selection and immortalized by stable integration of human telomerase as described [1]. They were maintained in EGM-2 with 5% fetal calf serum (FCS; GE Healthcare, Chalfont St Giles, UK) on surfaces coated with 2 mg/ml bovine fibronectin (Sigma-Aldrich, St. Louis, USA). LECs were used in passages 35 to 40.

Cell sorting and total RNA isolation

LECs were cultivated to a total number of around 7×10^7 cells. The cells were enzymatically detached, centrifuged at 100 ×*g* for 5 min and resuspended in cold EGM-2 to a concentration of 10×10^6 cells/700 ml. The mixed population was sorted with a MoFlo Astrios cell sorter (BD, Franklin Lakes, USA) according to the forward scatter (FSC) values. The cell suspensions were then centrifuged again at 100 ×*g* for 5 min and the medium supernatant was removed. The cells were resuspended in Trizol (Life Technologies, Carlsbad, USA) and chloroform (Carl Roth, Karlsruhe, Germany) was added. The suspension was mixed gently, left resting for 5 min at RT and afterwards centrifuged at 12,000 ×*g* for 15 min at 4 C. The RNA was precipitated by isopropanol for 10 min at RT. After centrifugation at 12,000 ×*g* for 15 min at 4 C, the RNA pellet was washed with 70% ethanol, dried at RT and resuspended in sterile water. Total RNA quality was estimated from 28S and 18S ribosomal RNA peaks on a Bioanalyzer 2100 instrument using the RNA 6000 Nano Kit (Agilent Technologies, Santa Clara, CA).

Standard transcriptome analysis

Isolated RNA from 3 technical replicates was used to produce biotinylated cRNA using the GeneChip HT 3' IVT Express Kit. Purified and

* Corresponding author.
E-mail address: wolfgang.holnthoner@trauma.lbg.ac.at (W. Holnthoner).

http://dx.doi.org/10.1016/j.gdata.2015.04.005

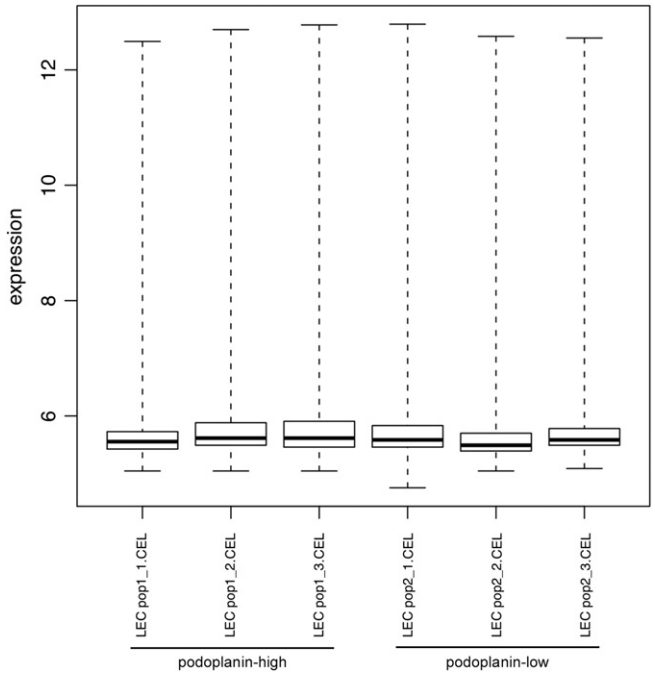

Fig. 1. Boxplots of the raw intensities.

fragmented cRNA was hybridized to GeneChip Human Genome U133 Plus 2.0 arrays (Affymetrix, SC, CA) following the manufacturer's recommendations. The Affymetrix GeneChip Fluidics Station 450 was used to wash and stain the arrays with streptavidin-phycoerythrin according to the standard protocol for eukaryotic targets (IHC kit, Affymetrix). Arrays were scanned with an Affymetrix GeneChip Scanner 3000. The resulting .CEL files were analyzed and normalized with Carmaweb (https://carmaweb.genome.tugraz.at/carma/). The raw data files were normalized using the robust multi-array average method (RMA) (Figs. 1 & 2). Raw and RMA normalized array data were submitted to Gene Expression Omnibus (GEO) and are available under the accession number GSE62510.

Enhanced transcriptome analysis to exclude false positive transcripts

As described [2], additional steps were taken to enrich for high quality data for the final selection of a set of differentially expressed genes. Using Carmaweb, a moderated t-test (limma) was performed on the

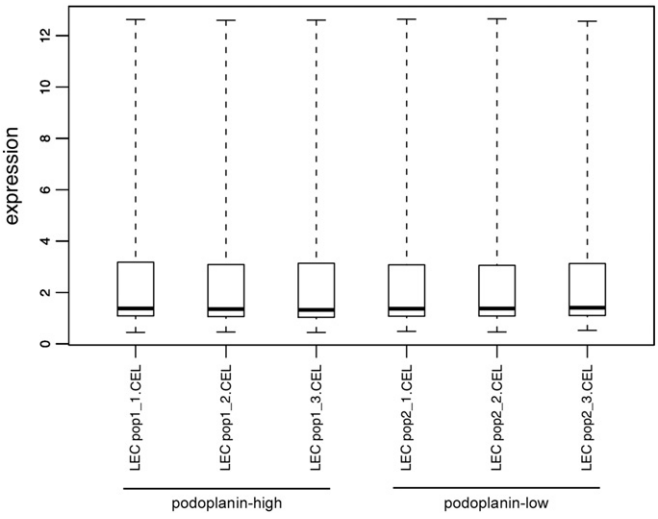

Fig. 2. Boxplots of the preprocessed expression values on each chip.

Table 1
RMA normalization *versus* MAS5 normalization, 100 genes with lowest raw p-values shown for both methods. Genes are ranked from lower to higher p-values for each given normalization method. For annotated genes, HGNC gene symbols are shown, else Affymetrix probeset IDs (xxxxx…_at) are given. Bold characters indicate genes common to both datasets.

Gene symbol or probeset ID, RMA normalized	Gene symbol or probeset ID, MAS5 normalized
1558048_x_at	SYTL4
234675_x_at	1570071_at
BMX	AKAP14
242881_x_at	LOC100289550
RASEF	LIMCH1
RASEF	239089_at
AFFX-r2-Bs-thr-5_s_at	PSMG4
SERPINE2	244791_at
224549_x_at	**RASEF**
M10098_3_at	FLT1
ANKRD11	BRD8
M10098_M_at	MMP7
NEAT1	232107_at
TMEM71	241618_at
HIP1	TCL1A
MGP	INPP4A
VWF	1569515_a_at
TPR	ZNF536
AFFX-r2-Bs-phe-M_at	CLIP1
PDPK1	1552955_at
M10098_5_at	RERE
AFFX-r2-Bs-dap-5_at	DYNC1H1
215626_at	C3
231199_at	UBE3B
BMP6	NLRP14
CDC27	**MGP**
RAD21	LOC650392
MDM4	GRM2
AFFX-DapX-5_at	FGF11
1565717_s_at	HMX2
230655_at	C21orf58
FMO3	C19orf21
PHACTR2	FOXP4
220038_at	243281_at
CAB39	C17orf52
SART3	1566042_at
239355_at	MTOR
RGS20	ZNF704
GPX3	CYP2A7
IDH3A	EVI5L
ITCH	MCM6
PPP1R3C	C3orf75
GATC	**SERPINE2**
FOXO3	hCG_1646157
P4HB	**GPX3**
PRUNE2	RANBP2
GPX3	SLC7A4
PICALM	**SCARB2**
SLC16A6	CIDEC
NEAT1	SH3GL1P2
AFFX-M27830_5_at	KLK8
PTER	CDK1
VEZF1	C11orf53
CPNE3	SNHG4
CD47	**P4HB**
SKIL	PAPPA
FNIP2	1564620_at
TMED2	242611_at
MARCH6	**RAD21**
ALDH1A1	NFKBIB
PALMD	HARBI1
ABCG1	**M10098_M_at**
ARHGAP18	FLJ10213
AFFX-r2-Bs-thr-M_s_at	**224549_x_at**
HTR2B	ARHGEF1
PAPPA	233687_s_at
SERPINE1	227223_at
MBNL1	RFFL
ZNF207	1558670_at

Table 1 (*continued*)

Gene symbol or probeset ID, RMA normalized	Gene symbol or probeset ID, MAS5 normalized
KSR2	SAMD1
ATP6V0E1	HORMAD2
AFFX-r2-Bs-phe-5_at	238796_at
SCD5	DEAF1
FMO3	KLF2
ZNF638	CDK1
SCARB2	222524_s_at
SNAP23	242276_at
AFFX-ThrX-M_at	**FMO3**
241773_at	LOC100130998
AFFX-r2-Bs-lys-5_at	C19orf34
DIRAS3	235355_at
COQ2	230750_at
242787_at	C14orf118
SMAD7	LOC643201
ARF6	RYK
MGEA5	207047_s_at
DLC1	HSP90B1
LOC100190986	OR1Q1
CXADR	PRIM2
S100A10	TNK2
IFITM1	SLC7A11
MSI2	208451_s_at
ID2	ONECUT3
SFRS6	**ZNF207**
AFFX-r2-Bs-dap-M_at	LIN7C
JAG1	**CPNE3**
LPCAT2	P2RX2
SBNO1	231005_at
MAT2A	GIMAP6
ANKHD1	234860_at

RMA normalized datasets, restricted to the 40% of the probesets with the biggest variance over all samples. To exclude potential normalization specific artifacts, a distinct normalization method, MAS5, values scaled to 200, was applied to the .CEL files, and differentially expressed genes were again determined by the moderated t-test (limma) on the normalized datasets, restricted to the 40% of the probesets with the biggest variance over all samples, again in Carmaweb. Depending on these two normalization methods, two distinct datasets for the 100 best candidates (100 lowest p-values) were generated (Table 1). These were combined and further analyzed in Microsoft Excel.

When screened for maximal differential gene expression in the podoplanin^{high} and podoplanin^{low} populations, the two normalization methods (RMA and MAS5) resulted in different top candidate lists. Only 12 transcripts of the 100 transcripts per list were commonly found in both lists. They are indicated in Table 1 by bold characters. Of the best 20 (lowest p-values) RMA-normalized genes, only 5 (25%) were also found in the 100 most regulated MAS5 normalized genes, while only 1 (5%) gene of the best 20 MAS5 normalized genes was found amongst the 100 most regulated RMA normalized genes. These differences raised concerns about selecting high numbers of false positive candidates by either normalization method.

To rule out this potential high number of false positive differentially regulated genes, an average of meanM (log2 transformed fold difference) and the respective statistic analyses (raw p-values, Bonferroni adjusted p-value — strong control of the family wise error rate), BH (Benjamini and Hochberg — strong control of the false discovery rate) was calculated from the combined lists. Table 2 depicts a ranking of the combined dataset by their meanM, averaged from both datasets. To select the most differentially regulated genes, the criteria for the means from both datasets were a raw p-value of <0.05, a BH value of <0.5, and a Bonferroni of <1 and only genes more than 2-fold regulated were selected (average meanM >1 AND < −1). From the remaining 40 genes, 10 found to be inversely regulated when comparing the different normalization methods were excluded as false positives, and additionally 5 internal Affymetrix probe-sets were excluded from

Table 2
MeanM based ranking of combined lists (100 genes each with lowest p-value for RMA normalization and MAS5 normalization, meanM values averaged). For annotated genes, HGNC gene symbols are shown, else Affymetrix probeset IDs (xxxxx…_at) are given.

Gene symbol or probeset ID	meanM average	Raw p RMA	Raw p MAS5
1558048_x_at	−3.58	1.13E-25	2.15E-02
SYTL4	−2.26	3.26E-16	1.66E-05
SNHG4	−2.19	5.13E-14	1.82E-03
UBE3B	−2.09	9.12E-14	3.90E-04
238796_at	−2.09	2.44E-12	2.41E-03
MMP7	−2.09	2.57E-12	1.45E-04
1552955_at	−2.08	1.16E-11	2.90E-04
DYNC1H1	−2.07	3.14E-11	3.11E-04
INPP4A	−2.04	3.45E-11	2.14E-04
NLRP14	−1.98	9.15E-11	4.13E-04
RASEF	−1.91	1.93E-10	1.01E-04
ANKRD11	−1.82	2.18E-10	2.05E-01
C17orf52	−1.75	2.35E-10	8.46E-04
224549_x_at	−1.74	6.10E-10	2.17E-03
234675_x_at	−1.70	1.07E-09	1.73E-02
FLT1	−1.67	1.13E-09	1.28E-04
OR1Q1	−1.66	1.34E-09	2.99E-03
242276_at	−1.65	1.69E-09	2.47E-03
227223_at	−1.63	3.16E-09	2.25E-03
1570071_at	−1.62	5.97E-09	2.60E-05
ZNF638	−1.61	1.29E-08	2.03E-01
NEAT1	−1.56	1.40E-08	7.25E-02
BRD8	−1.52	2.01E-08	1.38E-04
1566042_at	−1.50	3.02E-08	8.71E-04
215626_at	−1.48	3.17E-08	1.79E-01
FGF11	−1.44	4.78E-08	5.88E-04
232107_at	−1.41	7.11E-08	1.61E-04
RASEF	−1.41	8.14E-08	3.25E-02
207047_s_at	−1.36	1.40E-07	2.89E-03
242881_x_at	−1.34	1.60E-07	1.27E-02
PDPK1	−1.33	1.72E-07	8.48E-02
M10098_M_at	−1.31	2.04E-07	2.09E-03
M10098_5_at	−1.30	2.06E-07	4.04E-02
AFFX-r2-Bs-thr-5_s_at	−1.28	2.55E-07	1.05E-02
ITCH	−1.28	2.83E-07	3.66E-02
M10098_3_at	−1.27	3.09E-07	1.49E-02
TMEM71	−1.27	4.95E-07	5.18E-03
NEAT1	−1.21	7.36E-07	4.53E-02
1558670_at	−1.20	7.91E-07	2.30E-03
TCL1A	−1.19	9.28E-07	1.68E-04
230750_at	−1.18	1.04E-06	2.61E-03
231199_at	−1.15	1.06E-06	7.31E-02
HIP1	−1.15	1.08E-06	2.53E-02
RAD21	−1.14	1.16E-06	1.92E-03
AFFX-r2-Bs-dap-5_at	−1.12	1.31E-06	7.41E-03
1565717_s_at	−1.11	1.33E-06	7.37E-02
SART3	−1.11	1.42E-06	2.05E-02
C11orf53	−1.08	2.35E-06	1.74E-03
C3orf75	−1.07	2.38E-06	1.29E-03
GATC	−1.06	2.68E-06	1.09E-01
FOXP4	−1.06	2.69E-06	7.63E-04
CD47	−1.02	2.86E-06	5.20E-02
KLF2	−1.01	3.80E-06	2.44E-03
ATP6V0E1	−1.00	4.23E-06	5.23E-02
AFFX-DapX-5_at	−0.99	4.56E-06	2.44E-02
CPNE3	−0.99	5.00E-06	3.30E-03
AFFX-r2-Bs-phe-M_at	−0.99	5.14E-06	1.35E-01
SCARB2	−0.98	5.45E-06	1.47E-03
MARCH6	−0.98	7.30E-06	6.71E-02
RANBP2	−0.98	9.07E-06	1.42E-03
CDC27	−0.97	9.56E-06	8.96E-02
MDM4	−0.96	1.01E-05	2.95E-01
MBNL1	−0.95	1.02E-05	6.66E-03
FNIP2	−0.95	1.13E-05	8.85E-02
ZNF207	−0.93	1.15E-05	3.23E-03
TMED2	−0.93	1.19E-05	3.00E-02
230655_at	−0.93	1.26E-05	8.48E-02
VEZF1	−0.92	1.48E-05	1.72E-01
ARF6	−0.92	1.51E-05	7.04E-03
P4HB	−0.91	1.72E-05	1.84E-03
SFRS6	−0.91	2.10E-05	8.20E-03
RGS20	−0.89	2.14E-05	1.69E-01

(*continued on next page*)

Table 2 (*continued*)

Gene symbol or probeset ID	meanM average	Raw p RMA	Raw p MAS5
CAB39	−0.89	2.50E-05	1.56E-01
SNAP23	−0.89	2.86E-05	1.29E-02
AFFX-r2-Bs-thr-M_s_at	−0.89	3.01E-05	4.15E-02
PHACTR2	−0.89	3.35E-05	1.07E-01
FOXO3	−0.86	3.40E-05	2.33E-01
NFKBIB	−0.85	3.45E-05	1.93E-03
JAG1	−0.85	3.01E-05	3.02E-03
IDH3A	−0.84	2.04E-07	2.88E-03
RYK	−0.84	7.30E-04	3.29E-03
SMAD7	−0.84	1.13E-05	1.42E-03
ID2	−0.84	2.14E-05	1.66E-05
AFFX-ThrX-M_at	−0.84	7.30E-06	2.23E-04
AFFX-M27830_5_at	−0.83	7.91E-07	2.47E-03
SERPINE1	−0.83	2.68E-06	9.90E-05
MSI2	−0.82	2.10E-05	2.25E-03
MAT2A	−0.82	3.40E-05	3.11E-04
ARHGAP18	−0.81	2.35E-06	1.82E-03
TPR	−0.81	6.10E-10	1.38E-04
DLC1	−0.81	1.26E-05	3.19E-03
AFFX-r2-Bs-lys-5_at	−0.80	9.56E-06	6.68E-04
241773_at	−0.80	9.07E-06	1.19E-03
AFFX-r2-Bs-dap-M_at	−0.80	2.86E-05	2.29E-03
SBNO1	−0.80	3.35E-05	1.29E-03
PRIM2	−0.79	1.80E-04	3.90E-04
LIN7C	−0.77	3.37E-03	1.61E-04
PTER	−0.75	9.28E-07	2.41E-03
LOC100190986	−0.75	1.48E-05	2.42E-03
SCD5	−0.74	4.56E-06	1.28E-04
CXADR	−0.74	1.51E-05	2.44E-03
222524_s_at	−0.73	2.73E-02	2.99E-03
AFFX-r2-Bs-phe-5_at	−0.71	4.23E-06	2.89E-03
220038_at	−0.71	8.14E-08	1.74E-03
MGEA5	−0.70	1.19E-05	4.13E-04
SLC7A11	−0.70	5.51E-02	2.47E-03
SKIL	−0.70	1.16E-06	2.14E-04
COQ2	−0.68	1.01E-05	7.63E-04
DEAF1	−0.66	3.48E-01	1.45E-04
242787_at	−0.65	1.02E-05	1.86E-03
S100A10	−0.63	1.72E-05	6.26E-04
ANKHD1	−0.61	3.45E-05	2.30E-03
HARBI1	−0.39	1.24E-02	1.68E-04
PICALM	−0.34	4.95E-07	3.27E-05
C14orf118	−0.34	1.66E-01	2.55E-03
GIMAP6	0.62	5.47E-03	8.46E-04
SAMD1	0.64	5.54E-02	2.84E-03
239355_at	0.73	1.65E-07	9.59E-05
ARHGEF1	0.76	1.50E-01	3.36E-03
CDK1	0.77	2.10E-04	2.90E-04
PRUNE2	0.80	4.50E-07	7.42E-04
CDK1	0.80	1.00E-04	5.88E-04
242611_at	0.81	6.20E-03	1.45E-03
PPP1R3C	0.84	2.08E-07	2.96E-04
KSR2	0.84	3.11E-06	1.14E-03
DIRAS3	0.85	9.99E-06	1.02E-03
PAPPA	0.86	2.41E-06	3.47E-04
IFITM1	0.87	1.88E-05	1.93E-03
LPCAT2	0.92	3.28E-05	1.62E-03
PALMD	0.92	2.05E-06	3.37E-03
BMP6	0.93	6.02E-09	2.22E-03
208451_s_at	0.94	3.29E-03	1.38E-03
SLC16A6	0.95	4.98E-07	1.65E-04
ABCG1	0.96	2.17E-06	3.42E-03
C21orf58	0.96	9.76E-02	2.44E-04
CYP2A7	0.98	1.06E-01	1.28E-03
FMO3	0.99	4.85E-06	3.14E-03
HTR2B	0.99	2.40E-06	3.22E-03
CIDEC	1.01	1.63E-02	1.62E-03
GPX3	1.04	1.86E-07	5.55E-04
ALDH1A1	1.04	1.54E-06	2.49E-03
RFFL	1.07	1.06E-01	2.74E-03
EVI5L	1.10	1.30E-03	2.20E-03
GPX3	1.11	4.54E-07	2.27E-04
VWF	1.12	4.81E-10	2.60E-05
HSP90B1	1.15	6.99E-02	7.65E-05
C19orf21	1.15	5.88E-01	5.44E-04
MGP	1.23	3.44E-10	8.51E-05

Table 2 (*continued*)

Gene symbol or probeset ID	meanM average	Raw p RMA	Raw p MAS5
FMO3	1.27	5.74E-08	1.86E-03
hCG_1646157	1.27	2.42E-01	7.43E-05
SERPINE2	1.28	1.13E-11	2.58E-03
C19orf34	1.30	8.36E-01	2.40E-03
LOC100130998	1.36	1.76E-01	2.94E-03
C3	1.36	3.63E-01	2.31E-03
234860_at	1.43	2.29E-01	8.71E-04
BMX	1.52	3.55E-14	7.72E-04
1569515_a_at	1.53	1.70E-02	2.15E-03
231005_at	1.53	2.47E-01	1.53E-03
LOC650392	1.62	1.04E-01	1.97E-03
LOC643201	1.68	7.50E-01	1.87E-03
MCM6	1.69	2.18E-01	3.37E-03
1564620_at	1.74	7.50E-01	3.21E-03
P2RX2	1.77	6.47E-01	1.20E-03
ONECUT3	1.77	1.91E-01	1.62E-03
TNK2	1.78	1.96E-01	2.45E-03
HMX2	1.82	3.28E-05	6.26E-04
233687_s_at	1.82	1.88E-05	2.22E-03
244791_at	1.86	9.99E-06	9.90E-05
KLK8	1.94	4.85E-06	1.62E-03
PAPPA	1.95	3.11E-06	1.86E-03
GRM2	1.97	2.41E-06	5.55E-04
241618_at	1.97	2.40E-06	1.65E-04
PSMG4	1.98	2.17E-06	9.59E-05
MTOR	2.01	2.05E-06	1.02E-03
ZNF536	2.03	1.54E-06	2.27E-04
239089_at	2.07	9.71E-02	2.61E-03
SLC7A4	2.08	4.98E-07	1.45E-03
LIMCH1	2.11	4.54E-07	7.65E-05
HORMAD2	2.13	4.50E-07	2.40E-03
ZNF704	2.20	2.08E-07	1.14E-03
SH3GL1P2	2.29	1.86E-07	1.62E-03
CLIP1	2.29	1.65E-07	2.44E-04
LOC100289550	2.34	5.74E-08	7.43E-05
235355_at	2.34	6.02E-09	2.58E-03
RERE	2.35	4.81E-10	2.96E-04
FLJ10213	2.36	3.44E-10	2.15E-03
AKAP14	2.53	1.13E-11	3.27E-05
243281_at	2.56	3.55E-14	7.72E-04

the final list. This list, as published in [2] contains 25 more than two-fold differentially regulated transcripts.

This study was funded by the EU Biodesign Program (262948) and a Femtech student fellowship from the Austrian research promotion agency FFG.

References

[1] M. Groger, H. Niederleithner, D. Kerjaschki, P. Petzelbauer, A previously unknown dermal blood vessel phenotype in skin inflammation. J. Invest. Dermatol. 127 (12) (2007) 2893–2900.

[2] S. Rohringer, W. Holnthoner, M. Hackl, A.M. Weihs, D. Runzler, S. Skalicky, M. Karbiener, M. Scheideler, J. Proll, C. Gabriel, B. Schweighofer, M. Groger, A. Spittler, J. Grillari, H. Redl, Molecular and cellular effects of in vitro shockwave treatment on lymphatic endothelial cells. PLoS One 9 (12) (2014) e114806.