A transcriptomic signature predicting septic outcome in patients undergoing autologous stem cell transplantation

Yasmine Labiad, Geoffroy Venton, Farnault Laure, Céline Baier, Julien Colle, Cédric Mercier, Ivanov Vadim, Corinne Nicolino-Brunet, Loriod Béatrice, Nicolas Fernandez-Nunez, et al.

*Title Page

A transcriptomic signature predicting septic outcome in patients undergoing autologous stem cell transplantation

Running title: Predictive transcriptomic signature of sepsis

Labiad Yasmine [1], Venton Geoffroy [1,2*], Farnault Laure [2], Baier Céline [1], Colle Julien [1,2], Mercier Cedric [2], Ivanov Vadim [2], Nicolino Corinne [2], Loriod Béatrice [1], Fernandez-Nunez Nicolas [1],Torres Magali [1], Mattei Jean-Camille [3], Rihet Pascal [2], Nguyen Catherine [1] and Costello Régis [1,2 #]*

[1] Aix Marseille Univ, INSERM, TAGC, Marseille, France, [2] AP-HM, Service d'Hématologie et de Thérapie Cellulaire de La Conception, Marseille, France, [3]Aix Marseille Univ, INSERM, CRO2, Marseille, France

[*] *Equal contribution*

Paper's word count: 3174

 # Corresponding author :

Prof. Costello Régis (MD., PhD.), AP-HM Centre Hospitalier Universitaire La Conception, Service d'Hématologie et Thérapie Cellulaire, 147 Boulevard Baille, 13005 Marseille, France. Phone number: +33 4 91 38 41 50, Fax number: +33 4 91 36 41 54, E-mail address: regis.costello@free.fr

Complete contact informations

Labiad Yasmine *PhD* : Aix-Marseille Université - INSERM, UMR1090, TAGC Campus de Luminy, 163 Avenue de Luminy, case 928, Marseille, 13288, France Phone : +33 4 91 82 87 26 Mail : labiad.yasmine@gmail.com

Venton Geoffroy *MD-PhD*: Campus de Luminy, 163 Avenue de Luminy, case 928 Marseille, 13288. AP-HM Centre Hospitalier Universitaire La Conception, Service d'Hématologie et Thérapie Cellulaire, 147 Boulevard Baille, 13005 Marseille, France. Phone : + 33 4 91 38 41 50 Mail : geoffroy.venton@ap-hm.fr

Farnault Laure *MD*: AP-HM Centre Hospitalier Universitaire La Conception, Service d'Hématologie et Thérapie Cellulaire, 147 Boulevard Baille, 13005 Marseille, France. Phone : + 33 4 91 38 41 50 Mail : laure.farnault@ap-hm.fr

Baier Céline *PhD*: Aix-Marseille Université - INSERM, UMR1090, TAGC Campus de Luminy, 163 Avenue de Luminy, case 928, Marseille, 13288, France Phone : +33 4 91 82 87 26 Mail : celine.baier@yahoo.fr

Colle Julien *MD*: Aix-Marseille Université - INSERM, UMR1090, TAGC Campus de Luminy, 163 Avenue de Luminy, case 928, Marseille, 13288. AP-HM Centre Hospitalier Universitaire La Conception, Service d'Hématologie et Thérapie Cellulaire, 147 Boulevard Baille, 13005 Marseille, France. Phone : + 33 4 91 38 41 50 Mail : julien.colle@ap-hm.fr

Mercier Cédric *MD* : AP-HM Centre Hospitalier Universitaire La Conception, Service d'Hématologie et Thérapie Cellulaire, 147 Boulevard Baille, 13005 Marseille, France. Phone : + 33 4 91 38 41 50 Mail : cedric.mercier@ap-hm.fr

Ivanov Vadim *MD*: AP-HM Centre Hospitalier Universitaire La Conception, Service d'Hématologie et Thérapie Cellulaire, 147 Boulevard Baille, 13005 Marseille, France. Phone : + 33 4 91 38 41 50 Mail : vadim.ivanov@ap-hm.fr

Nicolino Corinne *MD*: AP-HM Centre Hospitalier Universitaire La Conception, Service d'Hématologie et Thérapie Cellulaire, 147 Boulevard Baille, 13005 Marseille, France. Phone : + 33 4 91 38 41 50 Mail : corinne.nicolino@ap-hm.fr

Loriod Béatrice : Aix-Marseille Université - INSERM, UMR1090, TAGC Campus de Luminy, 163 Avenue de Luminy, case 928, Marseille, 13288, France Phone : +33 4 91 82 87 26 Mail : beatrice.loriod@inserm.fr

Fernandez-Nunez Nicolas : Aix-Marseille Université - INSERM, UMR1090, TAGC Campus de Luminy, 163 Avenue de Luminy, case 928, Marseille, 13288, France Phone : +33 4 91 82 87 26. Mail : nicolas.fernandez-nunez@inserm.fr

Torres Magali : Aix-Marseille Université - INSERM, UMR1090, TAGC Campus de Luminy, 163 Avenue de Luminy, case 928, Marseille, 13288, France Phone : +33 4 91 82 87 26. Mail : magali.torres@univ-amu.fr

Mattei Jean-Camille: Aix-Marseille Université - INSERM, UMR 911, CRO2, Faculté de Médecine - 27, Boulevard Jean Moulin, Marseille, 13385, France. Phone: +33 4 91 38 00 00 Mail : jeancamille.mattei@gmail.com

Rihet Pascal *Pr*: Aix-Marseille Université - INSERM, UMR1090, TAGC Campus de Luminy, 163 Avenue de Luminy, case 928, Marseille, 13288, France Phone : +33 4 91 82 87 26. Mail : pascal.rihet@univ-amu.fr

Nguyen Catherine *PhD*: Aix-Marseille Université - INSERM, UMR1090, TAGC Campus de Luminy, 163 Avenue de Luminy, case 928, Marseille, 13288, France Phone : +33 4 91 82 87 26. Mail : catherine.nguyen@inserm.fr

Costello Régis *Pr*: Aix-Marseille Université - INSERM, UMR1090, TAGC Campus de Luminy, 163 Avenue de Luminy, case 928, Marseille, 13288, France. Service d'hématologie et de thérapie cellulaire APHM, Hôpital de la Conception, Marseille, France. Phone: + 33 4 91 38 41 50 Mail: regis.costello@free.fr

Highlights :

1) A transcriptomic signature predicts infection in neutropenic patients
2) Expression of eleven genes can be used routinely for sepsis prediction
3) Early anti-infectious treatment may improve sepsis outcome
4) Identification of a predictive transcriptomic signature is ongoing in leukemia

Abstract

Autologous hematopoietic stem cell transplantation (auto-HSCT) is a standard treatment in multiple myeloma and relapsed or refractory lymphomas. After auto-HSCT, hematologic reconstitution and infectious complications are the main two critical issues. Though many patients develop infectious complications after therapeutic intensification, it remains impossible to predict infection for each individual. The goal of this work was to determine and identify a predictive transcriptomic signature of systemic inflammatory response syndrome (SIRS) and/or sepsis in patients receiving auto-HSCT.

High throughput transcriptomic and bioinformatics analysis were performed to analyze gene expression modulation in peripheral blood mononuclear cells (PBMCs) in 21 patients undergoing auto-HSCT for hematological malignancies (lymphoma or multiple myeloma [MM]).

Transcriptomic analysis of PBMCs samples collected just after conditioning regimen identified an eleven genes signature (*CHAT, CNN3, ANKRD42, LOC100505725, EDAR, GPAT2, ENST00000390425, MTRM8, C6orf192 and LOC10289230 and XLOC-005643*) that was able to early predict (at least 2 to 7 days before its occurrence) the development of SIRS or sepsis.

The possibility of SIRS or sepsis occurrence early prediction (2–7 days before occurrence) opens up to new therapeutic strategies based on pre-emptive antibiotic and/or antifungal prophylaxis adapted to the specific risk profile of each patient.

Key words: Systemic Inflammatory Response Syndrome (SIRS); Sepsis; Autologous Hematopoietic Stem Cell Transplantation (Auto-HSCT); Transcriptomic Analysis; Treatment related mortality (TRM)

Abstract word count: 195

Introduction

Auto-HSCT is based on the administration of myelosuppressive high-dose chemotherapy, followed by infusion of autologous hematopoietic stem cells to obtain hematologic reconstitution. Hematopoietic stem cells (HSCs) infusion reduces chemotherapy-induced myelosuppression period and procedure-related mortality rate below 3% [1–3]. With few exceptions (solid tumors, autoimmune diseases), auto-HSCT is essentially indicated for selected hematological malignancies and considered as a standard treatment in young patients with MM and for relapsed or refractory lymphoma.

Besides direct toxicity of conditioning regimens, deep (<0.5 G/L neutrophils) and prolonged (usually 7–12 days) neutropenia exposes patients to significant risks of infection. The saprophytic gram-negative bacilli (such as *Escherichia coli*) are the most common cause of septic shock [4] and chronic immunosuppression exposes to the risk of fungal infection. An anti-fungal prophylaxis is usually administered [5] but antimicrobial prophylaxis is less often given because its effectiveness is not clearly established and increases the risk of *Clostridium difficile* diarrhea [6]. Unfortunately, it is not possible to foresee which patients will develop a SIRS and/or sepsis. Therefore, it remains impossible to adjust the antibiotic or antifungal prophylaxis to the specific risk profile of each patient.

The main objective of this work was to determine and identify a predictive transcriptomic signature of the SIRS and/or sepsis in patients receiving auto-HSCT, leading to the possibility of a pre-emptive anti-infectious treatment.

Patients, materials and methods

Patients

The prospective study was approved by the institutional review board of the Assistance Publique des Hôpitaux de Marseille (AP-HM – AORC2012 2012-08). Written informed consent was obtained from each patient. Patients were admitted in the hematology department of the Conception university hospital for undergoing auto-HSCT. All patients were under 65 years and were already followed in the hematology department for MM or high-grade lymphoma. Before auto-HSCT, patients were in complete remission (CR) or in partial remission (PR) after conventional chemotherapy. Inclusion criteria were the same as required for being eligible to auto-HSCT. Twenty fours patients were included in this study protocol. Twenty one were analyzed. Blood samples were collected for each patient at three moments: before the conditioning regimen (T1), after the conditioning regimen and before the graft infusion (T2) and at the end of the neutropenic phase (T3).

All patients (after written informed consent) hospitalized in the Hematology and Cellular Therapy Department of the university hospital of Marseille between August 2012 and December 2015 have been included. Only the patients (21 in all) in whom the RNA extraction was a success (after quantification and quality control) have been analyzed.

SIRS and sepsis definitions

SIRS and sepsis definition are based on the American College of Chest Physicians society of Critical Care Medicine consensus. SIRS is defined as the systemic inflammatory response to a variety of severe clinical symptoms with at least two of the following criteria: a) Temperature higher than 38 °C or lower than 36 °C b) Heart rate higher than 90 beats/min c) Respiratory rate higher than 20 breaths/min or PaCO2 lower than 32 mmHg d) White blood cell counts higher than 12,000 cells/mm^3 or lower than 4,000 cells/mm^3, or the presence of more than 10% immature neutrophils. The last criterion cannot be considered in auto-HSCT context because of the aplasia phase following HSC

infusion [7, 8]. Sepsis is defined as SIRS secondary to documented or suspected infection. Patients with severe sepsis are patients with sepsis and at least one organ dysfunction. Septic choc is defined by severe sepsis associated with refractory hypotension [8].

Conditioning regimens

Conditioning regimen for therapeutic intensification was high dose melphalan (200 mg/m^2) for patients with MM and BEAM for patients with lymphoma (carmustin 300 mg/m2 at day -6, etoposide 150 mg/m2 from day -5 to -2 twice daily, cytarabin 200mg/m2 from day -5 to -2 twice daily and melphalan 140mg/m2 day -1, with auto-HSCT on day 0).

Hematopoietic stem cell collection

HSCs mobilization required hematopoietic growth factors (G-CSF in most cases). Two procedures have been used to mobilize HSCs. For steady-state collection, G-CSF was injected at the dose of 10µg/kg-a-day. Mobilization chemotherapy consisted in high doses cyclophosphamide (1.5 to 4g/m^2) followed by 5µg/kg-a-day of G-CSF. When HSCs collection failed with these usual procedures, plerixafor (SDF-1/CXCR4 interaction inhibitor) was used at 240 µg/kg.

PBMCs isolation

Blood samples were centrifuged to separate plasma and other blood components. Concentrate blood was diluted and PBMCs were collected by Ficoll-Hystopaque density-gradient centrifugation [9].

RNA extraction, quantification quality controls

Depending on cell number, two kits were used for RNA extraction, the RNeasy Mini Kit Qiagen ™ (Qiagen, Valencia, California), which accepts 10 million of cells and the RNeasy Midi Kit Qiagen ™, which accepts up to 100 million cells. RNAs were extracted according to the Qiagen protocol. RNAs

were quantified by NanoDrop 1000 (Nano Drop Technologies, San Diego, CA). Optical density was measured at 260 and 280 nm and the ratio 260/280 (> 1.8) indicates its purity. Extracted RNA quality was checked with Agilent 2100 Bioanalyzer ™ (Agilent Technologies, Santa Clara, California). A score on a scale of 0 to 10 was automatically attributed to each sample and corresponded to RNA Integrity Number (RIN). Sample with a RIN under 7 was discarded.

Pangenomic gene expression assay

One hundred nanograms of total RNA was labelled using One-Color Microarray-Based Gene Expression Analysis: Low Input Quick Amp (LIQA) labelling protocol. 0.6 µg of the purified Cy3 labelled cRNA were hybridized for 17h at 65 °C, at 60 rpm, using the SurePrint G3 human GE 8x60K V2 chips (Agilent Technologies, Santa Clara, California). Microarrays were composed of 62 928 features. Probes synthesized on chips had a size of 60 nucleotides. Microarrays were washed using Gene Expression Wash Buffer Kit (Agilent Technologies) and scanned through the standard Agilent protocol. Data were processed using Feature Extraction software.

Differential gene expression analysis

The library AgiND is implemented in R software in order to analyze and visualize data. AgiND was developed on Bioconductor library model (tagc.univ-mrs.fr/ComputationalBiology/AgiND/) and is used to diagnose data quality and data-microarray normalization. Quantile method was used to normalize data; the objective was to homogenize distribution of microarray intensity [10]. A filter was applied on row data to delete controls, then a second filter was applied to delete genes which were expressed under the background in at least 80% of samples in each group (SIRS – , SIRS+, Sepsis+).

To test co-factors effects (gender, treatment, infection) on gene expression, GeneANOVA software was used to perform ANalysis Of Variance (ANOVA) on normalized data to determine an estimation

of the contribution of each factor (genes, gender, treatment, infection) in gene expression variation. Global ANOVA model is given in the following formula: $Y = \mu + \beta_1 G + \beta_2 T + \beta_3 I + \varepsilon$, where Y is explained variable, μ is global mean, β_1, β_2, β_3, are model coefficients, and G, T, I are the quantitative variables; $\beta_1 G$, is the gene effects, $\beta_2 T$, is the treatment effects, $\beta_3 I$, the infection effects, and ε is the error term [11]. Differential gene expression analysis was performed using linear Models for Microarray Data (Limma). Limma is a multivariate analysis and takes into account co-factors-effect tested by ANOVA analysis (treatment and gender) (**Table 2**). In order to determine infection impact on gene expression and on the heat map, new gene expression was calculated after subtracting gender and treatment co-factors effects. Unsupervised hierarchical clustering was applied on adjusted gene expression median adjusted data to group genes and samples, according to their expression using 'TMeV' (Tigr MultiExperiment Viewer) MeV: MultiExperiment Viewer | Part of the TM4 Microarray Software Suite [http://www.tm4.org/mev/]. Pearson correlation was used, and clusters were grouped on the basis of average linkage method.

Support Vector Machines (SVMs)

SVMs were applied to predict classification of patients according to the predictive signature. SVMs use training set in which genes known to be related to each other by function or samples related to a group are described as positive examples and genes or samples known not to be members of that class are labelled as negative examples. In this study, samples were attributed to two groups: SIRS and sepsis patients in the first group and patients with no temperature in the second. They were combined into a set of training examples used by SVM to distinguish class members from non-members on the basis of expression data. After learning the class expression features, the SVM can be used to recognize and classify each sample on the basis of their expression [11, 12]. SVMs tool is implemented in TMeV software MeV: MultiExperiment Viewer | Part of the TM4 Microarray Software Suite. [http://www.tm4.org/mev/].

High-throughput quantitative PCR

RT-PCR was performed using SuperScript™ VILO™ MasterMix protocol described by Invitrogen. Sixty nanogram of RNA of each sample was retrotranscribed, using 4µL of SuperScript™ VILO™ MasterMix, and RNase DNase Free water for a final volume of 20µL. Mix was incubated 10 min at 25 °C, 60min at 42 °C and 5min at 85°C. Quantitative PCR was performed with FlexSix array on BioMark™ HD (Fluidigm). A pre-amplification of each sample was applied according to the protocol provided by Fluidigm. The pre-amplification step was followed by an Exonuclease I treatment (BioLegend) to remove unincorporated primers. The final product was diluted 5-fold using 18µl of TE Buffer (10 mM Tris-HCl, 1mM EDTA). In a second part, 12 x 12 (samples x primers) qRT-PCR reactions are performed for each primer pair on each sample on the 12 x 12 array (FlexSix). We used the EvaGreen detection assay for following standard Fluidigm protocols. Ct values were calculated from the system software Biomark Real-Time PCR Analysis (Fluidigm). The primers used were:

Gene	R	F
SLC18B1	5'-GCAGGACAGCTTTTCAGTATCAC-3'	5'-CCTGGCCTTAAACATCACCG-3'
CNN3	5'-GGCTGGCTCCTTTATTAGTGC-3'	5'-AAGCTGGCCAAAGTGTAATTG-3'
MTMR8	5'-ACTTAAGTGGGGAGAGGGGT-3'	5'-GAGGCAGACTACTCCAAGCA-3'
GPAT2	5'-GCCCAGAGAAGCCTACATCA-3'	5'-CATCAGAAGCTCCTGGGGGA-3'
PLCG1	5'-ACAGGAATCTTGGTGCTTCAGT-3'	5'-CAGGGAGGTACATGGCCAAT-3'
ANKRD42	5'-CGCCCATGAAAAGCTGCATA-3'	5'-GGAGCAAATCTGACAGCCCA-3'
GAPDH	5'-CCACCACCCTGTTGCTGTGA-3'	5'-CCCACTCCTCCACCTTTGAC-3'
EDAR	5'-ACATGAGCTGACACTGGCTG-3'	5'-TGAGTGTGCCATTCCAGGAT-3'
TRAV3 -	5'-GCCTCGGTTGGGGTATTGAA-3'	5'-CTCAGCCGGAAGATCAGGTC-3'

CHAT	5’-CAGCAGAACATCTCCGTCGT-3’	5’-ATGGCCATTGACAACCACCT-3’
LOC100289230	5’-GTGCTTCCGGAAAACGTAAA-3’	5’-TCCTGTGCCCGTAATTTCTC-3’

157

Results

Patient's analysis

Twenty-four patients were included in the study, among these patients; 21 validated the molecular criteria for transcriptomic analysis. Among these 21 patients, 6 patients did not develop a fever or SIRS (28.6%), nine patients developed SIRS (42.6%), 5 a sepsis (23.8%) and 1 a severe sepsis (5%). Patients' clinical data are summarized in the **Table 1.**

Pangenomic gene expression

The 21 samples were analyzed on Agilent microarrays 'SurePrint G3 GE 8x60K human.' After 17 hours of hybridization, the chips were washed and scanned. Results passed microarrays quality controls. Raw data were transformed into *log2* and normalized with quantile method. 24 046 probes expressing a higher signal than background (in at least one group) have been selected. Two methods were used for statistical analysis: ANOVA analysis to measure the impact of each factor on gene expression variation (infection, gender, treatment) and Limma analysis on three groups (SIRS –, SIRS+, sepsis +) with the following co-factors: gender and treatment, to define differential gene expression **(Figure 1)**. ANOVA analysis was performed on the 24 046 probes to estimate impact of each factor (infection, gender, treatment) in gene expression variation. P-value was calculated for each factor. For all factors (genes, infections, gender, and treatment) p-values were $<10^{-4}$ **(Table 2)**. Infection F score was the highest, suggesting that 'infection' had the greatest impact on the gene expression. Genes F score was of 115.45 (p-value $<10^{-4}$) and explains genes fluctuations. Treatment and gender had an impact on the variation of gene expression with a F score of respectively 176.21 and 156.07 (p-value $<10^{-4}$). Unsupervised hierarchical clustering method was used to classify the differently expressed genes identified by Limma analysis. Expression similarity profiles of the genes were grouped on the horizontal axis and samples on the vertical axis. Gene expression profiles are

shown in heat map **(Figure 2)**. With a FDR (False Discovery Rate) fixed to 5%, 11 genes differentially expressed were identified between the patients who did not have temperature (SIRS-) and the patients who developed SIRS and sepsis (taking account effect of gender and treatment). The eleven differentially expressed genes were: *CHAT, CNN3, ANKRD42, LOC100505725, EDAR, GPAT2, ENST00000390425, MTRM8, LOC10289230* and *XLOC-005643* **(Table 3)**. Gene expression profiling distinguished two groups: patients who did not develop temperature, on the one hand, and patients with SIRS or sepsis, on the other hand. *CHAT, CNN3, C6orf192, ANKRD42, LOC100505725, EDAR, GPATZ, ENST00000390425, MTRM8, LOC10289230* were over expressed and *XLOC-005643* was under-expressed in patient samples with SIRS or Sepsis in contrast to patients who did not develop fever and SIRS.

SVM

SVM and leave-one-out cross-validation were used to classify patients according to their gene expression. 21 samples were separated into two groups, the first group was composed of 15 SIRS and septic patients and the second of the 6 patients who did not have temperature. SIRS and sepsis patients were considered as positive experiments, and the other as negative. All 15 patients of the positive experiments were classified as positive, staying in positive class (true positives=15) and none was transferred from negative class to positive (False negatives=0). On the 6 patients of the negative experiments, all were classified as negative, all retained in negative class (true negatives=15) and none was recruited into negative class from positives (false positives=0). We performed a Fisher exact test on our results; the p-value = 1.84 E-5.

RT-qPCR

In order to confirm the microarray gene expression results, RT-qPCR was performed on 9 genes that composed the transcriptomic signature. The *LOC 100289230* gene was differently expressed (p-value

= 0.003) in patients who developed SIRS or sepsis compared with patients who did not. *TRAV3, EDAR, PLCG1-AS1-001, GPAT2, MTRM8, CNN3,* and *SLC18B1* were also differentially expressed with p-values of 0.004, 0.01, 0.01, 0.005, 0.009, 0.09 and 0.01, respectively **(Figure 3)**.

Discussion and conclusion

This clinical study identified eleven genes significantly and differentially expressed in patients who developed SIRS or sepsis after the conditioning regimen for auto-HSCT. Ten of them were up-regulated (*CHAT, CNN3, ANKRD42, LOC100505725, EDAR, GPATZ, ENST00000390425, MTRM8, C6orf192 and LOC10289230*) while only one was down-regulated (*XLOC-005643*). All patients with this specific transcriptomic signature developed a SIRS or sepsis within 48 hours (range 48 hours-7 days) following conditioning regimen. All patients were classified in the right group according to their gene expression and based on SVMs analysis. After a wide scientific literature review, genes composing our predictive signature are not directly involved in sepsis or infection pathways. In this cohort of 21 patients, 9 developed a SIRS and 5 a sepsis. Only one patient developed a severe sepsis, and no patient had septic shock, thus impeding the possibility to identify a specific transcriptomic signature predictive for these life-threatening conditions. In addition, since a SIRS/sepsis predictive signature before the conditioning regimen was not identified, the transcriptomic signature was not linked to the patient pre-auto HSCT status but depended on the conditioning regimen patient's response.

We wondered about the definition of sepsis and SIRS and the clustering of sirs and sepsis; first, white blood cells count criteria is not relevant in patients in deep aplasia. Furthermore, in the same way, patients in deep aplasia present grade II-III anemia especially after BEAM conditioning regimen. In patients with no cardiovascular and/or pulmonary comorbidity, hemoglobin until 8g/dl is

231 tolerated. However, that often results in an increase of the heart rate higher than 90 beats/min
232 and/or respiratory rate higher than 20 breaths/min, especially in patients with fever.
233 Fever is neither sensitive or specific in conventional patients. But, in our severely
234 immunocompromised patients fever is more sensitive and specific than in immunocompetent
235 patients. Nevertheless, it is not unusual to start an antibiotic therapy in patients with no fever but
236 with a microbiological documentation and/or a major increase of the C reactive protein only.
237 At last, we have been very drastic on the definition of our sepsis patients group. Only the patients
238 with fever and a microbiological documentation have been considered in sepsis. Anyway, our
239 immunocompromised patients are paucisymtpomatic and very few infections were clinically
240 probable.

241 Our transcriptomic signature predicts SIRS/sepsis profiles and is more robust than the main
242 confounding factors, such as conditioning regimen, type or gender. BEAM conditioning regimen was
243 more myelosuppressive than melphalan alone and men have had a more significant risk to develop
244 infectious complications than women [13]. Auto-HSCT patients affected by MM are conditioned by
245 high dose of melphalan while patients with lymphoma by high dose BEAM chemotherapy, meaning
246 that treatment and pathology are confounding factors. SAM and ANOVA analyses were performed
247 on the data of patient samples before chemotherapy, and no differentially expressed genes were
248 found between patients affected by lymphoma or MM.

249 This work proposes a transcriptomic approach of the sepsis issue during auto-HSCT
250 neutropenic phase. Vanska et al. have shown that high pentraxin 3 level predicted septic shock and
251 bacteremia at the onset of febrile neutropenia after intensive chemotherapy of hematologic patients
252 [14]. Nonetheless, high pentraxin 3 level had only a predictive value for septic shock in patients who
253 already had a febrile neutropenia. In contrast, our predictive transcriptomic signature identifies
254 patients who have a major risk to develop SIRS and/or sepsis at least 48 hours (range 48 hours–7
255 days) before onset of fever. In order to confirm our results, our transcriptomic signature has been

validated on a 10 patient's prospective validation cohort (data not shown). Since total gene expression analysis by microarrays is not possible in clinical routine for time and cost-effective reasons, RT-qPCR analysis on a blood sample of the eleven deregulated genes will have to be developed in order to assess the possibility of a routine use.

Prophylactic anti-fungal and/or anti-microbial treatments have not proven a high efficiency regarding the outcome in patients undergoing highly myeloablative chemotherapy. For example, the use of ciprofloxacin induces a 20% reduction of infection rates but lead to 70% increase in the development of cifloxacin resistance [15]. Early identification of patients who will develop SIRS and/or sepsis could perhaps contribute to a better use of anti-infectious agents as preemptive treatment instead of a prophylactic treatment, although this is only an hypothesis by should be tested in a prospective clinical trial. A limitation of our study, directly related to the small number of patients and of infectious events, is the impossibility to identify a specific profile for microbial or fungal or viral infections, in order to precisely orientate the preemptive treatment. Finally, identification of high *vs* low infectious risk patients after myeloablative chemotherapy, could help to select patients with a safe septic profile and allows an early hospital discharge or even outpatient based blood stem cell transplantation, in order to develop a more rational utilization of hospital resources but still maintaining optimal safety conditions [16-19]. A validation of this transcriptomic signature in a prospective and larger cohort of patients is necessary to have a stronger clinical impact. Another strategy (work in progress) is to validate in RT-qPCR only the eleven genes of interest in a prospective and larger cohort.

Conflict of interest

All authors declare no financial conflict of interest

Funding

This work was supported by grants from AP-HM, Assistance Publique – Hôpitaux de Marseille (AORC APHM Junior), and Aix-Marseille Université, Advanced BioDesign. TGML platform supported by the France Génomique

Acknowledgments

We would like to thank INSERM, Aix-Marseille Université and Advanced BioDesign (Parc Technologique de Lyon Bâtiment Cèdre 1, 97 Allée Alexandre Borodine, 69 800 Saint-Priest – Lyon) for grants.

We are grateful to the patients who gave their informed consent to the use of their samples for research. We are grateful to the IBiSA Transcriptomics and Genomics Marseille-Luminy (TGML) platform was supported by the France Génomique National infrastructure. We thank Laurence Borge for assistance and the use of the cell culture platform facilities (CRCM U1068, Marseille) Plateforme de Culture Cellulaire de Marseille-Luminy, bâtiment TPR2 163 Avenue de Luminy, 13009 Marseille. We also thank Geneviève Victorero and Noushine Mossadegh-Keller for technical supports.

References

[1] Kurnick N. Autologous and Isologous Bone Marrow Storage and Infusion in the Treatment of Myelo-Suppression. Transfusion. 1962;2:178-187.

[2] MCFARLAND W, GRANVILLE NB, DAMESHEK W. Autologous bone marrow infusion as an adjunct in therapy of malignant disease. Blood. 1959;14:503-521.

[3] Clifford P, Clift R, Duff J. Nitrogen-mustard therapy combined with autologous marrow infusion. The Lancet. 1961;277:687-690.

[4] Veeraputhiran M, Jain T, Deol A, et al. BEAM Conditioning Regimen Has Higher Toxicity Compared With High-Dose Melphalan for Salvage Autologous Hematopoietic Stem Cell Transplantation in Multiple Myeloma. Clinical lymphoma, myeloma & leukemia. 2015;15:531-535.

[5] Maertens J, Marchetti O, Herbrecht R, et al. European guidelines for antifungal management in leukemia and hematopoietic stem cell transplant recipients: summary of the ECIL 3—2009 update. Bone marrow transplantation. 2011;46:709-718.

[6] Sohn BS, Yoon DH, Kim S, et al. The role of prophylactic antimicrobials during autologous stem cell transplantation: a single-center experience. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology. 2012;31:1653-1661.

[7] Physicians ACoC. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. Critical care medicine. 1992;20:864-874.

[8] Robertson CM, Coopersmith CM. The systemic inflammatory response syndrome. Microbes and Infection. 2006;8:1382-1389.

[9] English D, Andersen BR. Single-step separation of red blood cells, granulocytes and mononuclear leukocytes on discontinuous density gradients of Ficoll-Hypaque. Journal of immunological methods. 1974;5:249-252.

[10] Bolstad BM, Irizarry RA, Åstrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics. 2003;19:185-193.

[11] Brown MP, Grundy WN, Lin D, et al. Knowledge-based analysis of microarray gene expression data by using support vector machines. Proceedings of the National Academy of Sciences. 2000;97:262-267.

[12] Quackenbush J. Computational analysis of microarray data. Nat Rev Genet. 2001;2:418-427.

[13] Adrie C, Azoulay E, Francais A, et al. Influence of gender on the outcome of severe sepsis: a reappraisal. CHEST Journal. 2007;132:1786-1793.

[14] Vänskä M, Koivula I, Hämäläinen S, et al. High pentraxin 3 level predicts septic shock and bacteremia at the onset of febrile neutropenia after intensive chemotherapy of hematologic patients. Haematologica. 2011;96:1385-1389.

[15] Pohlen M, Marx J, Mellmann A, et al. Ciprofloxacin versus colistin prophylaxis during neutropenia in acute myeloid leukemia: two parallel patient cohorts treated in a single center. Haematologica. 2016:haematol. 2016.147934.

[16] Ferrara F, Palmieri S, Viola A, et al. Outpatient-based peripheral blood stem cell transplantation for patients with multiple myeloma. The hematology journal : the official journal of the European Haematology Association / EHA. 2004;5:222-226.

[17] Mank A, van der Lelie J, de Vos R, Kersten MJ. Safe early discharge for patients undergoing high dose chemotherapy with or without stem cell transplantation: a prospective analysis of clinical variables predictive for complications after treatment. Journal of clinical nursing. 2011;20:388-395.

[18] Faucher C, Soriano ALC, Esterni B, et al. Randomized study of early hospital discharge following autologous blood SCT: medical outcomes and hospital costs. Bone marrow transplantation. 2012;47:549-555.

[19] Martino M, Morabito F. Autologous stem cell transplantation in multiple myeloma is not dead but alive and well. Expert opinion on biological therapy. 2015;15:149-154.

348 ***Table legends:***

349 **Table 1**: **Patients, pathologies and auto-HSCT characteristics**

350 **Table 2: co-factors effect on gene expression.**

351 ANOVA analysis was performed to test the co-factors impact (infection, treatment and gender) on
352 gene expression for samples taken at T2 (after chemotherapy). With p-values <0, 05 infection,
353 treatment and gender have a significant effect on the gene expression. DF: Degrees of Freedom

354 **Table 3: Differentially expressed gene characteristics**

355 Ch: Chromosome, +1: Forward strand, from 3' to 5', -1: Reverse strand, from 5' to 3'

356

357

358 ***Figure legends:***

359 **Figure 1: Differential gene expression flow chart**

360 Pangenomic array for the transcriptomic analysis have been used, data were filtered and the controls
361 were suppressed, cofactors effects were tested based on their implication on variation of gene
362 expression. A multivariate analysis – LIMMA – using the cofactors already tested was performed to
363 define the gene differentially expressed using Bioconductor library. The LIMMA model is given in the
364 following formula: $Y = \alpha + \beta_1 . T + \beta_2 . G + \beta_3 . I + \varepsilon$, where Y is the explained variable, α is the global
365 mean, $\beta_1 . T$, $\beta_2 . G$, $\beta_3 . I$, are the model coefficients, T, G, I, are the quantitative variables, β_1 T, is the
366 treatment effect, $\beta_2 . G$, is the gene effect, $\beta_3 . I$ is the infection effect and ε is the error term. Then, we
367 performed multi-testing correction by fixing the threshold to 5%. Finally, we adjusted the expression
368 data based on the co-factors before performing the hierarchical clustering *: Linear Models for
369 Microarray data.

370

371 **Figure 2: Gene expression profile adjusted on the effects of the cofactors**

372 Microarray analysis revealed 11 genes differentially expressed. This genes cluster distinguished SIRS
373 or sepsis patients from patients who didn't develop any fever or SIRS.
374
375 **Figure 3: Transcriptomic validation signature by RT-qPCR**
376 Relative gene expression composed the transcriptomic signature of patients who developed SIRS or
377 sepsis compared to patients who did not develop fever and SIRS
378 * = 0.05; ** = 0.01; *** ≤0.01, NS = not significant
379

Table1

Table 1 : Patients, pathologies and auto-HSCT characteristics

Baseline and Demographic Characteristics	n = 21
Patients age (median)	58 years Range [32-68]
Patient gender	
Male	14 (67%)
Female	7 (33%)
Hematologic malignancy	
Lymphoma	2
Multiple Myeloma	19
Conditioning Regimen	
BEAM	2
Melphalan high dose	19
Injected CD 34^+ (x 10^6 / kg)	2.82 Range [0.751-3.64]
Mean duration of fever	3 days SEM [-4.30 - 4.30]
Mean duration of neutropenia	9 days SEM [-2.90 – 2.90]
Documented bacterial infections	6 (30%)
Gram-positive bacteria	75%
Staphylococcus hominis	1
Staphylococcus epidermidis	2
Staphylococcus aureus	1
Streptococcus mitis	1
Clostridium difficile	1
Gram-negative bacteria	25%
Escherichia coli	1
Klebsiella pneumonia	1
Documented fungal infections	0 (0%)
Patient classification	
SIRS-	6 (29%)
SIRS+	9 (42%)
Sepsis+	5 (24%)
Severe sepsis+	1 (5%)
Septic shock+	0 (0%)

Table 2: Co-factors effect on gene expression.

ANOVA analysis was performed to test the co-factors impact (infection, treatment and gender) on genes expression for samples taken at T2 (after chemotherapy). With p-values < 0, 05 infection, treatment and gender have a significant effect on gene expression. DF: Degrees of Freedom

Factors	Sum of squares	DF	Variance	F score	P-value
Genes	3476905.49	28114	123.67	115.45	$< 10^{-4}$
Infection	954.39	1	954.39	890.96	$< 10^{-4}$
Treatment	188.75	1	188.75	176.21	$< 10^{-4}$
Gender	167.18	1	167.18	156.07	$< 10^{-4}$
Residual	602328.16	562297	1.07	-	-
Total	4080543.97	590414	6.91	-	-

Table 3: Differentially expressed gene characteristics.

Ch : Chromosome, +1 : Forward strand, from 3' to 5', -1 : Reverse strand, from 5' to 3'

Gene	Complete name	Ensembl ID	Location	Sens
SLC18B1 / C6ORF192	solute carrier family 18 member B1	ENSG00000146409	Chr 6: 132,769,370-132,798,553	-1
CNN3	calponin 3	ENSG00000117519	Chr 1: 94,896,949-94,927,278	-1
MTMR8	myotubularin related protein 8	ENSG00000102043	Chr X: 64,268,081-64,395,431	-1
LOC100505725 PLCG1-AS1-001 / TOP1-AS1 /	PLCG1 antisense RNA 1	ENSG00000226648	Chr 20: 41,098,329-41,138,003	-1
CHAT	choline O-acetyltransferase	ENSG00000070748	Chr 10: 49,609,095-49,665,104	1
EDAR	ectodysplasin A receptor	ENSG00000135960	Chr 2: 108,894,471-108,989,372	-1
ANKRD42	ankyrin repeat domain 42	ENSG00000137494	Chr 11: 83,193,739-83,260,694	1
GPAT2	glycerol-3-phosphate acyltransferase 2, mitochondrial	ENSG00000186281	Chr 2: 96,021,946-96,039,451	-1
ENST00000390425/ TRAV3	T cell receptor alpha variable 3 (gene/pseudogene)	ENSG00000211777	Chr 14: 21,723,713-21,724,321	1
LOC100289230	NS	NS	Chr 5: 98,929,134-98,931,009	1
XLOC_005643	lnc-CMAHP-1:1 / linc-FAM65B-1/RP3-425P12.2	ENSG00000230372	Chr 6: 25,061,853-25,063,735	-1

Figure1

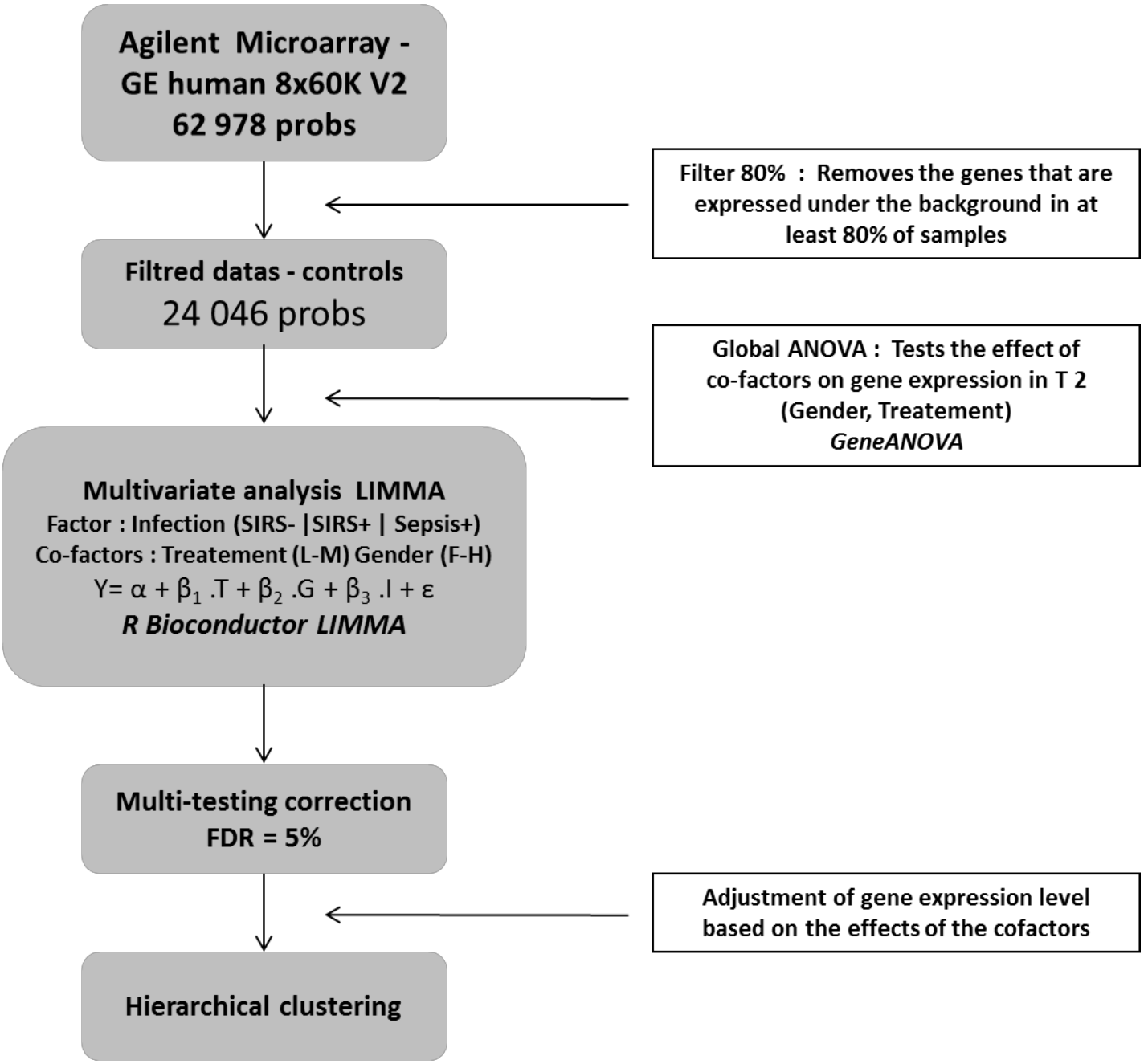

Figure 1: Differential genes expression flow-chart

Figure2

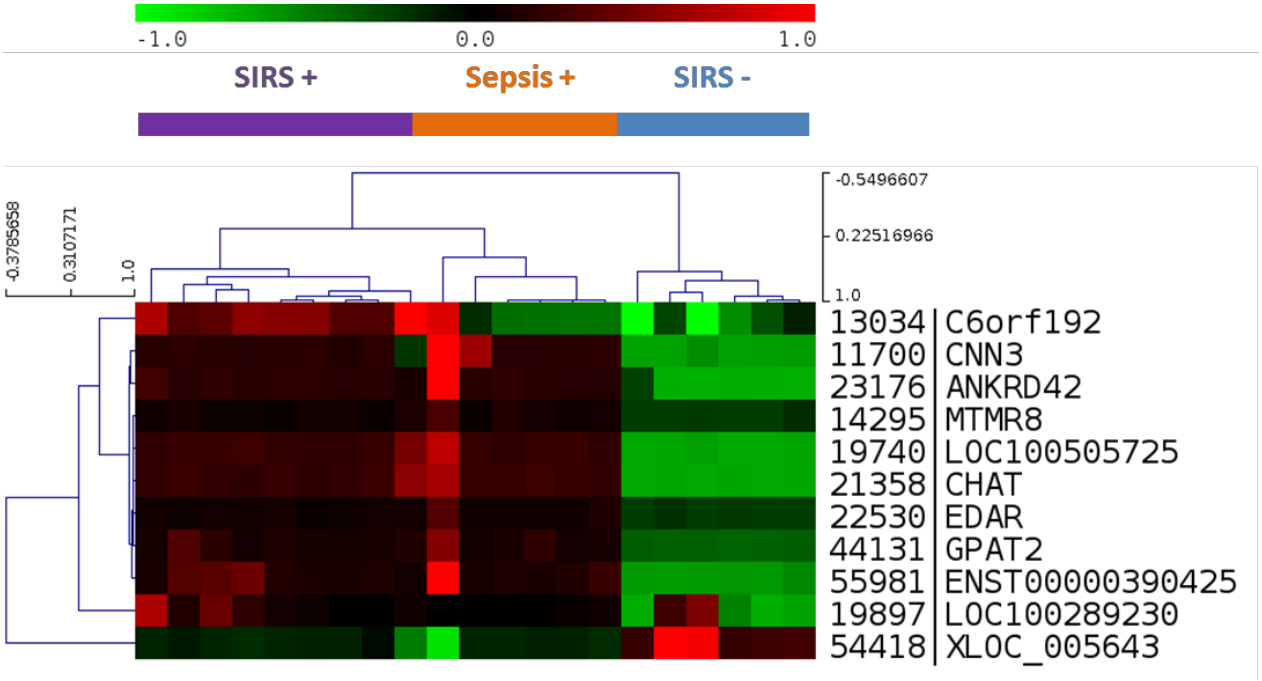

Figure 2: Gene expression profile adjusted on the effects of the cofactors

Color should be used

Figure3

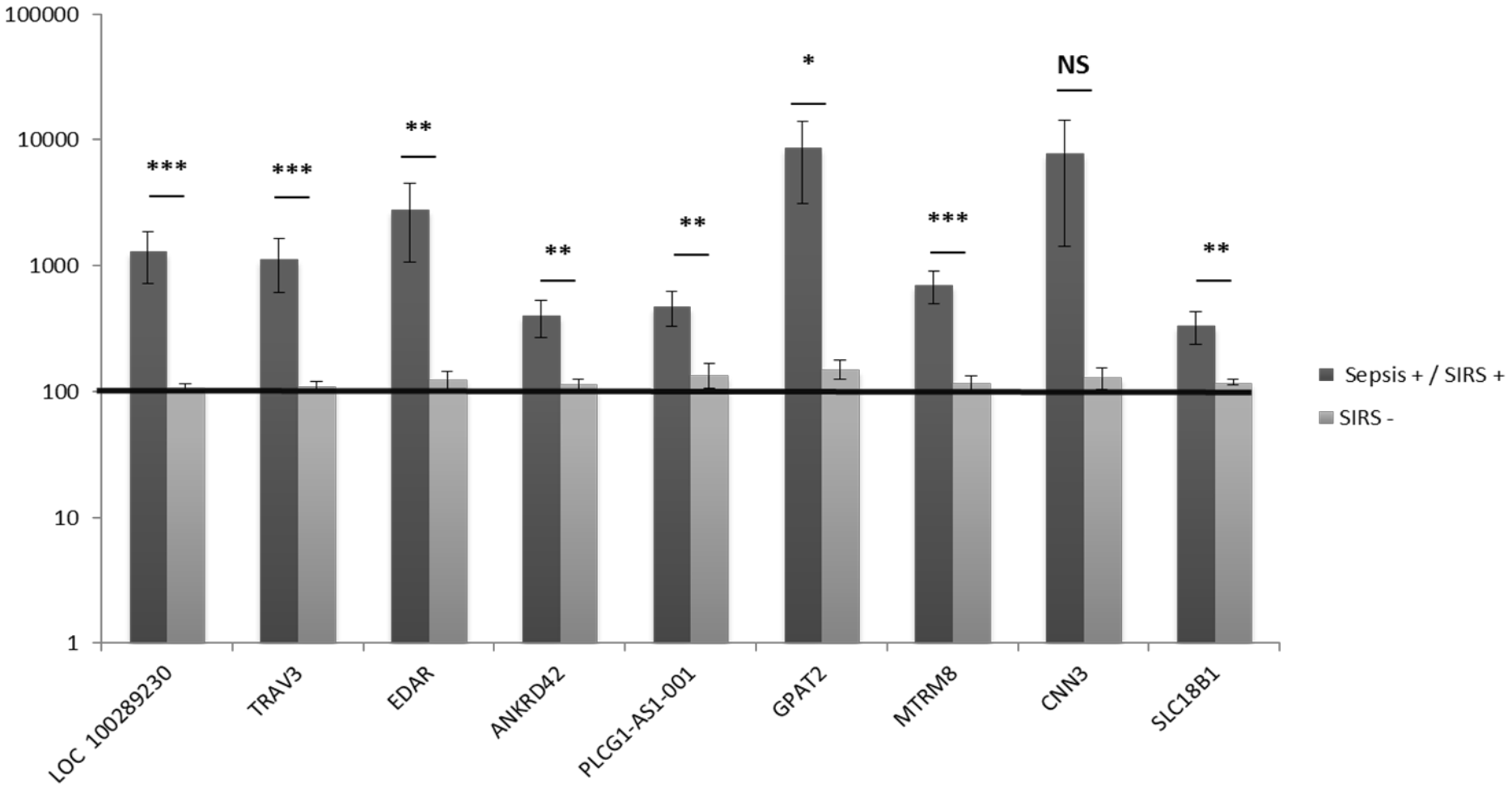

Figure 3 : Transcriptomic validation signature by RT-qPCR