UNIVERSITY OF TURIN

Department of Clinical and Biological Sciences

PhD Program in Experimental Medicine and Therapy

29th cycle

Transplantation induces a profound and persistent change in the transcriptional asset of human hematopoietic stem cells

PhD student: Petiti Jessica

Tutor: Prof. Daniela Cilloni

PhD coordinator: Prof. Giuseppe Saglio

A.A. 2017

SSD MED/15

INDEX

ABSTRACT **5**

INTRODUCTION **6**

Hematopoiesis 7

HSCs self-renewal regulation 8

Hematopoietic Stem Cell Transplantation 9

Indications for HSCT *10*

Transplant classification *11*

HSCs sources *12*

Procedural considerations *14*

BMT vs. UCBT *15*

UCB stem cells *ex* vivo expansion 16

The self-renewal potential after transplantation 17

AIM OF THE STUDY **19**

MATHERIALS & METHODS **21**

Cells source 22

Induced pluripotent stem cell (iPS) line maintenance 22

CD34+ cell enrichment 23

RQ-PCR analysis by Low density array/Microfluidic card 23

Immunofluorescence staining 24

Sparse Multivariate Analysis 25

Performance metrics 26

Network inference and Functional Analysis 27

NiCord – Gamida cells (UCB *ex vivo* expansion) 29

RESULTS **30**

Gene Expression Analysis 31

CD34+ cells after UCBT express genes of self-renewal, stem cell maintenance and cell reprogramming *31*

UCB vs adult HSC: self-renewal, stem cell maintenance and reprogramming genes are not differentially expressed *32*

Gene related to self-renewal, reprogramming and stem cell maintenance decline after transplantation with adult HSC *32*

Genes regulating self-renewal, cell reprogramming and stem cell maintenance are over-expressed in CD34+ cells after UCBT but not after adult HSCT *32*

Some reprogramming genes are similarly expressed in CD34+ cells after UCBT and iPS although their respective signature is divergent *33*

mRNA results are confirmed by protein analysis 36

Sparse multivariate regularization analysis 38

Only six genes are sufficient to discriminate UCB from adult HSC *39*

A signature of fifty-two genes discriminate cord blood CD34+ cells before and after transplantation *41*

A signature of sixty-two genes discriminate adult CD34+ cells before and after transplantation *44*

After transplantation adult versus cord blood CD34+ cells acquire a divergent transcriptional asset; reprogramming genes are a relevant part of such difference. *48*

CD34+ cells from UCB *ex vivo* expanded show a similar behavior to Adult post UCBT 51

Infusion procedure could affect stem genes expression 52

DISCUSSION **53**

REFERENCES **58**

SUPPLEMENTARY **63**

ABSTRACT

Hematopoietic Stem Cell Transplantation (HSCT) is a consolidated therapy in malignant and non-malignant disorders treatment. However, the biology of Hematopoietic Stem Cell (HSC) during and after the hematopoietic reconstitution remains largely unexplored. It has been argued since long time that HSC might undergo some sort of exhaustion after transplantation. Conversely, in children transplanted with Umbilical Cord Blood (UCB) the reconstitution seems to be much superior compared to that of children given adult HSCs. So, we used HSCT as in vivo model to investigate about many aspects of HSCs biology.

The expression of ninety-one genes, mainly involved in HSC regulation, were analyzed by quantitative Real-Time PCR (qRT-PCR) in CD34+ cells isolated from: Umbilical Cord Blood (UCB), Bone Marrow from healthy donors (BM), BM after UCB Transplant (UCBT), BM taken after transplantation with adult HSC. Protein expression was evaluated by Immunofluorescence for ten of these genes, selected after qRT-PCR analysis.

A multivariate learning algorithm identified four gene signatures shedding light on the molecular differences between the CD34+ cells groups. Transplanted CD34+ cells show a new transcriptional program completely changed from the cells of origin, however the cell source determines the definitive transcriptional asset. In particular, DPPA2, LIN28, NANOG, NES, OCT4, SOX1, SOX2 and PTEN were found unexpectedly over-expressed in CD34+ cells after UCBT but not after adult transplant. Thus, UCB and adult CD34+ cells use different molecular mechanism for accomplishing the hematopoietic reconstitution.

INTRODUCTION

Hematopoiesis

Hematopoiesis is the formation of blood cellular components that occurs during embryonic development and throughout adulthood to produce and replenish the blood system. Studying hematopoiesis can help us to better understand the processes behind blood disorders and cancers. Furthermore, hematopoietic stem cells (HSCs) can be used as a model system for understanding tissue stem cells and their role in ageing and oncogenesis [1]. In humans, hematopoiesis begins in the yolk sac and transitions into the liver temporarily before finally establishing definitive hematopoiesis in the bone marrow (BM) and thymus.

HSCs reside in the medulla of the bone and have the ability to give rise to myeloid or lymphoid progenitor cells [2]. Myeloid progenitors give rise to thrombocytes, erythrocytes, granulocytes and monocytes; while lymphoid progenitors give rise to Natural-Killer, T and B-lymphocytes (Fig. 1)

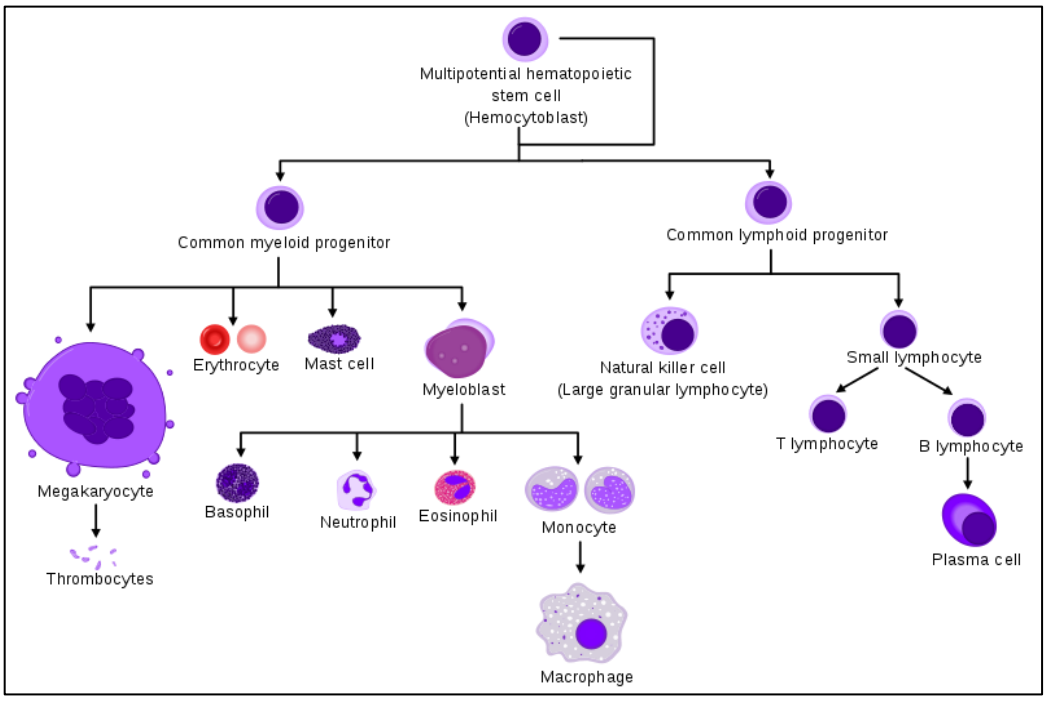


Figure 1: development of different blood cells from HSC to mature cells.

Red and white blood cells production is regulated with great precision and increasing rapidly in determinate condition, for exampleduring infection. The proliferation and self-renewal of HSCs depend on growth factors (GF). GF initiate signal transduction pathways, which lead to activation of transcription factors (TF). Mutations in TF are tightly associated with blood cancers, as acute myeloid leukemia or acute lymphoblastic leukemia.

HSCs self-renewal regulation

There are many factors and pathways that are important for HSCs self-renewal. Wnt e Notch signaling pathways were identified as the most important ones. The Wnt family of molecules is known to be crucial for embryonic development [4]. Most studies have found a positive role for Wnt in HSCs during development and regeneration. Activation of Notch signaling has been shown to promote HSCs expansion/self-renewal in both mice and humans in adult hematopoiesis [5].

Notch ligands exposure expands over 100-fold the populations of human cells expressing CD34, a cell surface protein marker for HSCs. Further studies demonstrated that when Notch ligand-expanded cord blood (CB) progenitors were used in a clinical setting, there was a rapid recovery of myeloid cells, indicating rapid engraftment of *ex vivo* expanded cells in humans [6]. These results indicate the importance of Notch-dependent regulation of hematopoietic development. Similarly, the microenvironment is known to be essential for the regulation and maturation of many stem cells. Several studies identify the osteoblast as an important cell that regulated the HSCs maturation in the BM [7, 8]. Furthermore,

there is evidence that vascular and stromal cells are also required for HSCs homeostasis [9].

Hematopoietic Stem Cell Transplantation

The hematopoietic stem cell transplantation (HSCT) is a prime example of successful applied regenerative medicine. For more than 30 years, HSCT has become a routine treatment for blood disorders and malignant diseases. A key element of its feasibility and success relies on the redundancy of HSCs in the BM. In fact, with 1% of donor BM cells it is possible to regenerate a new hematopoietic system within a relatively short period of time. After the eradication of the patient's own hematopoietic system, the transplanted donor hematopoietic stem and progenitor cells (HSPCs) provide lifelong reconstitution of the blood system of the patient.

The experimental evidence that HSCs naturally migrate back and forth from the BM periodically, as well as the identification of agents that increase HSCs mobilization (e.g., granulocyte colony-stimulating factor [G-CSF]), have opened new avenues for HSCT. However, although HSCT work successfully in the clinic, further improvement of the method is needed to minimize engraftment failure and post transplant infections. Actually a improvement for grafts with limiting numbers of HSCs (umbilical cord blood) and for gene therapy approaches for monogenetic inherited blood disorders is the *ex vivo* expansion of HSCs.. However, despite decades of research, the robust expansion or even maintenance of HSCs *ex vivo* is not yet routinely achieved [10].

Indications for HSCT

Hematological malignancy, such as acute myeloid leukemia, acute lymphoid leukemia, myelodisplastic and myeloproliferative syndromes, is the most common indication for HSCT in both children and adults. While the choice of HSCs source depends on both patient and disease characteristics, with malignant diseases, the speed of HSCs availability is often critically important. Over the years, the use of HSCT as a therapeutic modality has been extended to a variety of non-malignant hematological disorders, such as hemoglobinopathies, Fanconi anemia and metabolic storage diseases, and other neoplasm, for example breast cancer, glioma, microcellular lung cancer and some autoimmunological diseases [11] (Fig. 2).

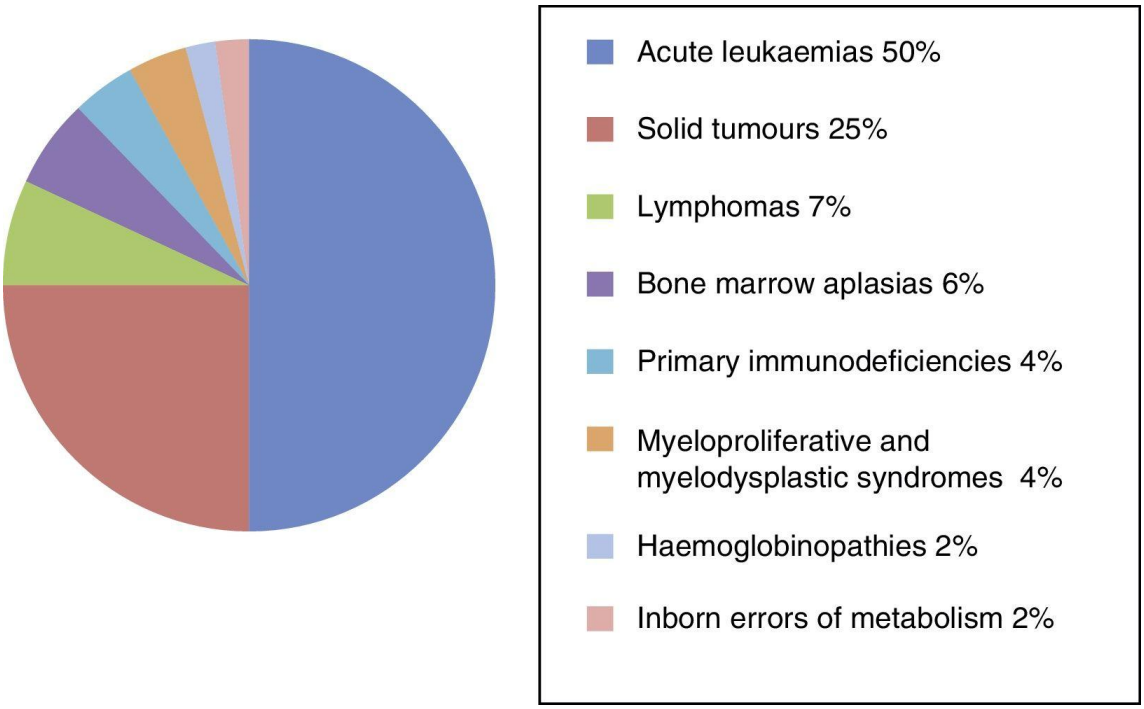


Figure 2: indications for HSCT

Transplant classification

HSCT are classified in:

- Autologous stem cell transplantation: stem cells are removed from a patient, stored, and later given back to that same person. Patients were treated with high doses of chemotherapy or a combination of chemotherapy and radiation. The high-dose treatment kills cancer cells, but also eliminates the blood-producing cells that are left in the BM. Afterward, the collected stem cells are transplanted back into the patient, allowing the BM to produce new blood cells.
- Allogeneic stem cell transplantation: stem cells were transferred from a healthy person (the donor) to patient after high-intensity chemotherapy or radiation. A donor is often a brother, sister or other relative. Unrelated donor HSCT requires allele level HLA matching between the donor and recipient. Considering that the current standard of 'suitable' HLA matching is more stringent than that in the past, finding an HLA matched donor has become more difficult in unrelated donor transplant settings. Public awareness, clinical/social network, and large registry databases may help to reduce this problem [12, 13]. In the absence of an HLA matched donor, CB can serve as an alternative HSCs source.

HSCs sources

Nearly two-thirds of patients requiring HSCT will not have a suitable related donor, so the applicability of HSCT to larger numbers of patients has been augmented with the increasing availability of unrelated donors. Currently, alternative HSCs sources include donor BM or peripheral blood stem cells (PBSC) and umbilical cord blood (UCB)[14].

BM harvesting

Although HSCs harvesting was performed for the first time more than 50 years ago, BM harvesting was developed mainly to perform allogeneic HSCT and later autologous transplantations. BM harvesting has become a relatively routine procedure. BM is generally aspirated from the posterior iliac crests while the donor is under either regional or general anesthesia. This can be a difficult procedure in donors who are smaller than the recipient, such as sibling donors, and several aspirations may be required for an adequate mononuclear cell (MNC) dose. The acceptable cell dose harvested in BM and required for allogeneic transplantation is $3–5\times10^8$ TNCs per kilogram of recipient body weight (BW)[15]. Complications related to BM harvesting are rare and involve anesthetic, infectious, and bleeding problems (Fig. 3).

PB HSC harvesting: mobilization and apheresis techniques

HSCs circulate in blood, albeit in very low concentrations, and can be identified and quantified using flow cytometry. It is necessary to mobilize HSCs from BM to PB.

Administration of recombinant hematopoietic growth factors (ie, the cytokines G-CSF and GM-CSF) to patients or donors down-regulates the adhesion molecules on the CD34+ cells and releases them into the PB, which can be collected by apheresis procedure [16]. Combined chemotherapies associated with the hematopoietic

growth factor currently used to mobilize HSCs from BM to PB in patients [17]. Different apheresis devices have been developed to harvest PB HSCs. All these techniques share a common process: separation of blood components in layers by centrifugation and harvest blood-mobilized HSCs in a particular phase associated with other blood cells. The number of HSCs to be harvested varies depending on the teams and nature of the transplantation, ie, from 4 to 10×10^6 CD34+ cells/kg recipients' BW. G-CSF commonly leads to side effects such as bone pains, malaise, headaches, chills, and (sometimes) fever (Fig. 3).

CB harvesting

After the birth of the infant, the umbilical cord is double-clamped from the umbilicus and transacted between the clamps. The umbilical cord vein is punctured under sterile conditions and the blood flows freely by gravity into an anticoagulated sterile closed harvesting system [18, 19] (Fig 3).

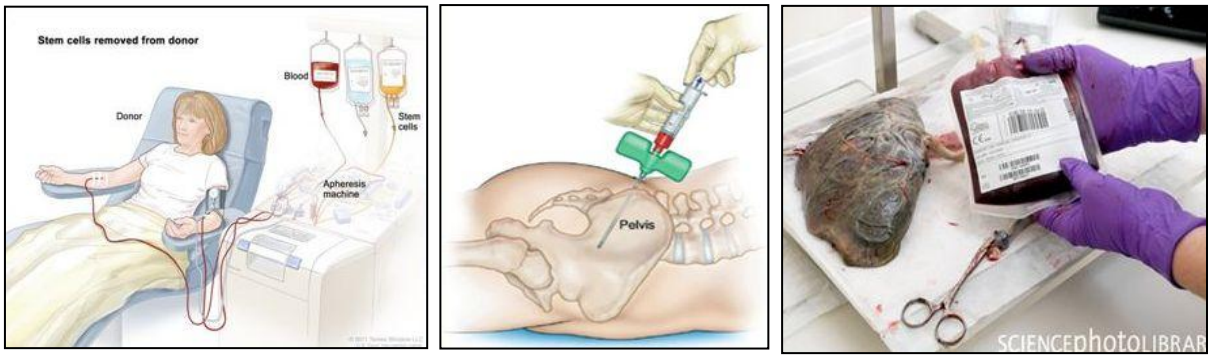


Figure 3: HSCs sources

Procedural considerations

Preparative or conditioning regimens involve delivery of maximally tolerated doses of multiple chemotherapeutic agents with no overlapping toxicities and may be classified as follows:

- Myeloablative regimens – These are designed to kill all residual cancer cells in autologous or allogenic transplantation and to cause immunosuppression for engraftment in allogeneic transplantation; they may be further sub classified as radiation-containing or non–radiation-containing;
- Nonmyeloablative regimens – These are immunosuppressive but not myeloablative and rely on the graft-versus-tumor effect to kill tumor cells with donor T cells.

Infusion of either BM or PB progenitor cells is a relatively simple process that is performed at the bedside. The BM product is infused through a central vein over a period of several hours. The HSCs engraft within the BM cavity by homing like mechanisms that have not yet been fully elucidated. Minimal toxicity is observed in most cases. AB0-mismatched BM infusions could occasionally lead to hemolytic reactions. Dimethylsulfoxide (DMSO), which is used for the cryopreservation of stem cells, may give rise to facial flushing, tickling sensation in the throat, and strong taste in the mouth (the taste of garlic).

After transplant, all patients are kept in high-efficiency particulate air–filtered, positive-air-pressure–sealed rooms, and strict hand hygiene is practiced. Patients who received an autograft may be managed in an outpatient setting, as they have a brief period of neutropenia and fungal infections. Most patients receive antibacterial and antifungal prophylaxis.

BMT vs. UCBT

In allogeneic HSCT for malignant disease, the therapeutic efficacy is attributed to both the cytotoxic effect of radio/chemotherapy and the anti-leukemic protection mediated by donor T cells, the so called Graft-versus-Leukemia effect (GvL) [21]. Immunological complications after HSCT, such as graft rejection and Graft-versus-Host-Disease (GvHD), are primarily caused by differences in histocompatibility antigens between donor and recipient. Thus, in HSCT with an HLA-mismatched donor, the risk of GvHD, graft failures and lethal opportunistic infections is increased. Therefore selection of a suitable donor with a match for the human leukocytes antigens (HLA) A, B and DRB1 on the allele level is critical [12, 22].

There are several advantages of using UCB over BM stem cells for transplants. The first advantage is that UCB is relatively easy to collect and process. After it is saved and sent to a storage facility, the UCB is quickly available for use within days to weeks after processing. In contrast, BM stem cells can take much longer to find a match, collect the sample, and process [23]. Furthermore, compared to UCB, BM collection and transplantation of stem cells are more costly [24, 25]. Another advantage of using UCB stem cells is the decreased risk of the transmission of infectious disease. This particular advantage is partly because UCB is almost never contaminated by Epstein-Barr virus or cytomegalovirus [25]. Considered to be immunologically immature, UCB stem cells produce significantly fewer natural killer cells, creating a substantial decrease in rejection. Consequently, UCB stem cells require less rigorous antigen tissue matching for transplants than BM stem cells [26].

Despite the benefits of using UCB stem cells for transplant, the process also has some disadvantages. For HSCT to be successful, measurable signs of engraftment

must occur. Two measurable signs are the recovery of both neutrophil and platelet production, that occur later in UCBT than in BMT [24, 27]. One of the factors that influence engraftment time is cell dose [28]. Because of the limited volume of cells collected from UCB, the amount of stem cells is approximately 10% less than BM [24]. This problem is greater for adults and adolescents, that need comparatively more quantity of stem cells for transplant rather than children. Additionally, it is unknown how long UCB will maintain its usefulness while frozen. Research indicates that UCB stem cells can be maintained up to 15 years, but it is unknown if the cells would be preserved over the entire lifetime of a person [27, 29]. Furthermore, there are the financial costs associated with maintaining the UCB over time.

UCB stem cells *ex* vivo expansion

The main drawback in UCBT is the low quantity of HSCs, since the content per UCB unit ranges between 0.4 and 1.0×10^9 total MNC, whereas the dose currently recommended ranges from 2.0×10^7 to 2.5×10^7 MNC/kg. A dose lower than 1.5×10^7 MNC/kg showed poor results [30], thus restricting UCB transplantation to pediatric patients in most of the cases. Since the biology of HSCs and their microenvironment are not totally understood [31], it has not been easy to overcome this issue. On this basis, *Ex vivo* expansion of HSCs from UCB and other sources became an alternative to increase the cell-dose available for transplants and to further research on HSCs. There is evidence that even if short-term expansion may modify HSC properties, it is strongly probable that the engraftment characteristics remain unaltered [32].

HSCs require an adequate microenvironment to keep their stem properties. In BM the microenvironment is very complex, HSCs are surrounded by bone matrix and different cells including fibroblast, adipocyte, macrophage and endothelial cells, which produce various cytokines and growth factors; these signaling molecules induce HSCs to differentiate or to remain in the stem state maintaining a balance in hematopoiesis. In vitro HSCs culture requires a suitable microenvironment, for that reason, different culture media, growth factors and supplements have been tested, but the optimal combination and concentration of growth factors to preserve the stem state has not been yet established.

The self-renewal potential after transplantation

To repopulate an adult BM, it is speculated that CD34+ cells expand itself approximately by a factor of 2 logs after BMT and 3 logs after UCBT. To achieve this result, HSCs must reorganize their transcriptional asset and probably activate self-renewal genes.

The sustained self-renewal potential of donor HSCs is critical for maintaining the long-term durability of the graft. HSCs are thought to be capable of self-regeneration *in vivo* over a lifetime without an apparent limit under homeostatic conditions [33, 34]. In contrast, it is well known that the repopulating ability of HSCs can be significantly compromised in transplant recipients [35, 36]. Several studies have demonstrated that the functional HSC units reach only 4% to 10% of normal levels after each transplantation [37, 38]. The self-renewal ability of HSCs is intrinsically limited [39], but extrinsic factors, such as the transplantation procedure [38] and the irradiated BM microenvironment [37], are also likely

involved. Despite the importance of this issue, the potential exhaustion of donor HSCs in transplant recipients has not been resolved at the molecular level. However, Iscove and Nawa [33] have elegantly disputed this concept. They showed that in children after UCBT the reconstitution of the HSCs reservoir (operationally LTC-IC) nearly approaches the *restitutio ad integrum*. In fact, it is much superior compared to that of children given adult HSCs (i.e., BM cells), notwithstanding both neutrophil and platelet recovery remains delayed [40]. Thus, UCB HSCs seem to display remarkable self-renewal machinery.

AIM OF THE STUDY

To investigate how HSC reorganize their transcriptional asset to cope with the need of BM repopulation, we evaluated the expression of 91 genes selected for their role in self-renewal and maintenance of stemness. CD34+ cells taken from patients transplanted with either adult or UCB HSCs were included to observe how these HSC sources respond to the transplant challenge.

In this study we examined the self-renewal program of HSCs of different origins by providing a transcriptional landscape of UCB and adult HSCs before and after transplantation by making use of different donor/recipient combinations. First, an exploratory analysis was performed by means of a standard Mann-Whitney hypothesis test, leading to a set of genes with significantly altered regulation. Then, a multivariate sparsity-inducing machine learning algorithm was used to identify four gene signatures with remarkably accurate predictive capabilities. The model predicting the considered donor/recipient combinations, was trained on a portion of the available samples and tested on the remaining previously unseen data. This approach evaluates the prediction performance in a statistically unbiased way. Furthermore, the four signatures underwent a functional characterization procedure that lead to a set of meaningful KEGG pathways as well as an inferred network of gene associations.

MATHERIALS & METHODS

Cells source

The study was approved by the ethic committee (number of approval 230/CEI). CD34+ cells were enriched from 12 UCB units and 15 adult donors (8 G-CSF-mobilized PB stem cells (PBSC) and 7 BM harvests). In addition, CD34+ cells were enriched from BM collected at different time points after HSCT in the following combinations of stem cell source/recipient: 13 adult patients transplanted with single UCB unit via intra-bone, 29 patients (5 adult and 24 pediatric) transplanted with adult HSCs. The clinical characteristics of patients are summarized in supplementary table 1. Additional CD34+ cells were separated from 10 UCB, 9 PBSC from adult donors, 15 BM of adults after UCBT, 10 BM samples from patients after HSCT from adult donor. These 44 samples were analyzed at the protein level only. Furthermore, CD34+ cells from CB *ex vivo* expanded (NiCord-Gamida cells) and from a patient transplanted with UCB *ex vivo* expanded (NiCord-Gamida cells) were collected.

Induced pluripotent stem cell (iPS) line maintenance

Human iPS cell lines (kindly provided by Dr Niels Geijsen, Hubrecht Institute for Developmental Biology and Stem Cell Research, Utrecht - Netherlands) were derived from human skin fibroblasts by transducing *Oct4, Klf4, Sox2, c-Myc* transgenes using lentiviral vectors as described [41, 42]. For expansion, iPS colonies were manually picked in a 4-7 day frequency and plated onto inactivated mouse embryonic fibroblast (MEF) feeder layer. The cells were cultured in

Dulbecco's modified Eagle's medium (DMEM)/F12 supplemented with 20% KnockOut Serum Replacement, 1% L-glutamine, 1% nonessential amino acids, 0.1 mM β-mercaptoethanol, 0.5% penicillin/0.5% streptomycin and 20ng/ml basis fibroblast growth factor (FGF)-D and incubated in a humidified tissue culture incubator containing 5% CO2 at 37°C. To prevent apoptosis, cells were treated with the ROCK inhibitor Y-27632 (10μM).

CD34+ cell enrichment

Mononuclear cells (MNC) were isolated from fresh BM, PBSC and CB by density gradient separation and enriched for CD34+ antigen by magnetic immune-selection method (Miltenyi Biotech, Berisch Gladbach, Germany), according to the manufacturer's protocol. Following two cycles on separate columns, CD34+ cells purity was assessed by flow cytometry: only samples with CD34+ cells ranging between 90% to 95% were selected for RNA extraction.

RQ-PCR analysis by Low density array/Microfluidic card

Total RNA was extracted from iPS and CD34+ cells using TRIzol and RNAqueous-Micro Kit (Ambion, Life Technologies, Carlsbad, CA), respectively, and cDNA was synthesized by Reverse Transcription PCR using High Capacity cDNA Reverse Transcription kit (Applied Biosystems, Life Technologies, Carlsbad, CA). All reactions were performed using T100 Thermal Cycler (Bio Rad, Hercules, CA). For the present study, 384 wells Taqman Low Density Array or MicroFluidic Card

(MFC) (Applied Biosystems, Life Technologies, Carlsbad, CA) have been used. Each well contains lyophilized specific primers and probes for respective target genes. The expression levels of ninety-one genes were investigated. A list of the examined genes and their main functions is provided in supplementary table 2. Among those genes, a discrete number is involved in many processes of stem cell self-renewal, stem cell maintenance, quiescence, cell cycle and somatic cell reprogramming. Each card also contains three different housekeeping genes as control: *GUSB*, *rRNA18S* and *ABL*. All the analysis has been performed in duplicate. Two samples have been analyzed in each card. RQ-PCR card data were analyzed by SDS2.3 (ABI prism 7900 HT Fast Real-Time PCR System, Life Technologies, Carlsbad, CA) software. Universal Human Reference RNA (cat. #740000-41, Stratagene, La Jolla, CA) was used as calibrator. Raw data were initially normalized against endogenous control gene (*GUSB*) and then compared to the calibrator.

Immunofluorescence staining

CD34+ cells were fixed with 4% Para-FormAldeide and permeabilized with 0.1% Triton X100, as previously described [44]. After blocking with 10% Fetal Bovine Serum + 5% BSA + 1% fish gelatine in PBS 1x for 1hr, cells were labeled using anti human antibodies at a final concentration of 4 μg/ml in blocking solution (Santa Cruz Biotechnology, Heidelberg, Germany). A complete list of the antibodies used is available in supplementary table 2. As secondary antibody Alexa Fluor 488-labelled goat anti-rabbit IgG (cat. A11008), goat anti-mouse IgG (cat. A11001) and donkey anti-goat IgG (cat. A11055) (Invitrogen Life Technologies, Carlsbad, CA) were used in a dilution of 1:500 in PBS 1x. Cells were then treated with Propidium

Iodide for 5 min to stain the nuclei and mounted with Mowiol. Cells were analyzed on a fluorescence microscope (Leica DM2000 LED, Leica microsystem, Germany) and images were captured using 40X/0.65 objective. Images were acquired using Leica application Suite 4.4.0 software and quantified using Image J, available at: http://rsbweb.nih.gov/ij/download.html.

Sparse Multivariate Analysis

l_1l_{2FS} is an embedded regularization method for variable selection capable to identify subsets of discriminative genes. The algorithm can be tuned to give a minimal set of discriminative genes or larger sets including correlated genes. The method is based on the elastic net optimization principle presented in Zou and Hasti [45] and further developed by De Mol et al [46, 47] and successfully applied in the analysis of molecular high-throughput data [48-51].

Assume we are given a collection of n samples, each represented by a d-dimensional vector x of measurements (e.g., the gene expressions), each sample is also associated with a binary label y, assigning it to a class (e.g. UCB or UCBT).

The dataset is therefore represented by a $n \times d$ matrix X, and Y is the n-dimensional labels vector. Using only a subset of the given data (training set), the l_1l_{2FS} algorithm looks for a linear function $f(x)=\beta^*x$, whose sign gives the classification rule that can be used to associate a sample to one of the two classes. The classification performance of $f(x)$ is then assessed on the remaining samples (test set) that were not used to build the model function. Note that the vector of weights β^* is forced to be a sparse vector, that is some of its entries are zero, then some variables (gene) will not contribute in building the estimator $f(x)$. The weight

vector β* is found in the so-called model selection phase, which consists in selecting the optimal values for two regularization parameters denoted with τ* and λ*, respectively. Model selection and classification accuracy assessment are performed within two nested K-fold cross-validation loops, in order to guarantee an unbiased result. As a consequence of the external loop of cross validation, $l_1 l_{2FS}$ provides a set of K lists of discriminant variables, therefore it is necessary to choose an appropriate criterion in order to assess a common list of relevant variables. We based ours on the absolute frequency, so we decided to promote as relevant variables the most stable genes across the lists. The threshold we used to select the final lists was chosen according to the slope variation of the number of selected genes vs. frequency, its value being 50%. In this way we managed to cut out those variables that were not stable across the cross-validation lists.

Performance metrics

We evaluated the prediction performance through the accuracy and the Matthews Correlation Coefficient (MCC) metrics. In customary notation, when considering a classification task, a classifier assigns the considered samples to two possible classes: negative (-1) and positive (+1). In this case, the true positives TP are the positive examples correctly classified ad +1, the true negatives TN are the negative examples classified as -1, the false negatives or Type II error FN are negative examples misclassified as members of the positive class and, similarly, the false positives or Type I error FP are the negative examples wrongfully assigned to class +1. Accuracy is used as a statistical measure of how well a binary classification test correctly identifies or excludes a condition. In other words, the accuracy is the

proportion of true results (both TP and TN) among the total number of cases examined. Accuracy ranges from 0% to 100%, the perfect classification. A random classifier would achieve an accuracy rate based on the prevalence of the two classes. If the prevalence is the same, i.e. the amount of samples is equal in the two classes, a random classifier would achieve 50% accuracy. MCC is a metric that takes into account all the parameters just defined above and it is defined as:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

MCC, unaffected by the presence of unbalanced classes, ranges between -1 and +1. The greater MCC the better the prediction with negative score marking below random performance.

Network inference and Functional Analysis

In order to verify if the identified genes belonging to the respective signatures were functionally associated, we used the Search Tool for Recurring Instances of Neighbouring Genes (STRING), a publicly available web-server able to find a set of potentially functionally associated genes to a gene query list [54]. The results of this functional characterization analysis, underline if the identified genes within each signature are connected through eight possible types of edge connections (Conserved neighborhood, Gene Fusions, Phylogenetic co-occurrence, Co-expression, Database imports, Large-scale experiments, Literature co-occurrence, Combined score), represented in different colors in the corresponding plots. For the functional analysis of the gene signatures we also used the on-line gene set

enrichment analysis toolkit WebGestalt [55]. The toolkit performs the functional characterization by a gene set enrichment analysis in several databases including Gene Ontology [56] and the Kyoto Encyclopedia of Genes and Genomes (KEGG) [57]. Given a KEGG pathway and a reference set (such as the entire human genome). The enrichment is based on the comparison between the fraction of signature genes in the pathway and the fraction of pathway genes in the reference set. The signature is enriched in the KEGG pathway if the former is larger than the latter fraction. To perform the enrichment analysis in KEGG, we selected the WebGestalt human genome as reference set, p–value ≤ 0.05 as level of significance, as the minimum number of genes and the default Hypergeometric test as statistical method.

NiCord – Gamida cells (UCB *ex vivo* expansion)

NiCord, based on the company's NAM technology, is in phase 3 clinical testing as a universal BM transplant solution. NiCord is an ex vivo expanded cell graft derived from CB stem cells. It is in development to be a curative treatment for more than 50% of high risk leukemia and lymphoma patients who need a BMT but who do not have the required fully matched donor. NiCord achieved the first evidence of *an ex vivo* expanded graft providing a significant, long-lasting hematopoietic recovery following transplantation. NiCord is designed to be smoothly integrated into current transplant protocols. Once designated for transplantation with NiCord, an eligible UCB is rapidly selected, and the manufacturing process, spanning approximately three weeks, begins. During this time, the patient receives a standard conditioning regimen lasting approximately eight to eleven days (Fig. 4).

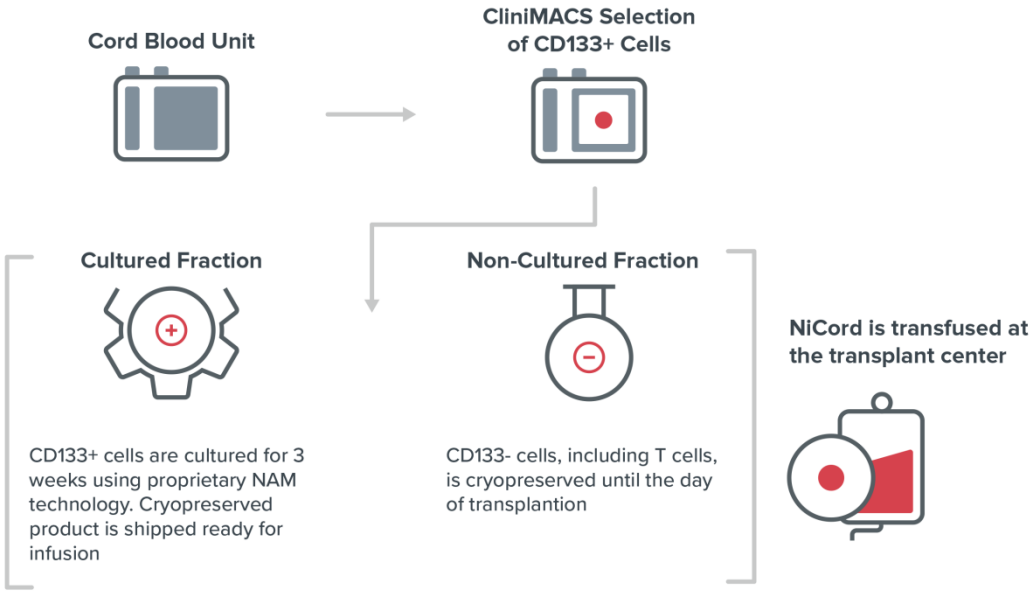


Figure 4: NiCord strategy for CB stem cells *ex vivo* expansion

RESULTS

Gene Expression Analysis

The expression of 91 genes was measured in CD34+ cells enriched from 12 UCB units, 15 adult donors and from BM collected after HSCT in 13 adult patients transplanted with UCB and in 29 patients transplanted with adult HSCs. The heterogenous population of samples did not prevent the identification of very reproducible gene expression clusters characterizing specific cells populations. The pair wise comparison among groups allowed to identify the most relevant genes and pathways discriminating them. We observed several gene signatures by means of differential expression and multivariate regularization analysis. For the former, we used standard Mann-Whitney test, while for the latter we used $l_1 l_{2\mathrm{FS}}$, and set in a nested cross-validation structure to ensure reproducibility and robustness. We pooled together both mobilized PBSC and BM stem cells, referring to the pooled group as adult HSC as no significant differences appeared when comparing the two subgroups.

CD34+ cells after UCBT express genes of self-renewal, stem cell maintenance and cell reprogramming

To validate the hypothesis that self-renewal of HSC after UCBT was more efficient than after adult HSC transplant [40], we started to analyze genes expressed in CD34+ cells after UCBT compared to CD34+ UCB cells. Among 91 genes analyzed, Mann-Whitney test allowed to identify 10 top genes differentially expressed. In detail, *DPPA2, LIN28, NANOG, NES, OCT4, PTEN, SOX1, SOX2*, were significantly up-regulated in CD34+ cells after UCBT compared to CD34+ cells obtained from UCB Units ($p<0,01$; Figure 5). We observed a wide variation on the expression of these genes in CD34+ cells after UCBT among samples. Altogether, there was statistical

difference with respect to the UCB CD34+ cells. Most of these genes play a key role in reprogramming somatic cells and are used in different combinations to generate iPS starting from somatic cells [41, 58, 59] (supplementary table 2). By contrast, *HOXB3* and *HOXB4* appeared down-regulated. As shown in Figure 5, we further extended the analysis by performing a comparison between: UCB vs. adult HSC, adult HSC vs. adult after HSCT, and adult after UCBT vs. adult after HSCT.

UCB vs adult HSC: self-renewal, stem cell maintenance and reprogramming genes are not differentially expressed

The pattern of expression of *LIN28, NANOG NES, OCT4 PTEN SOX1* and *SOX2* was similar in UCB and adult HSC. Only the expression level of *DPPA2, HOXB3* and *HOXB4* was significantly decreased in adult HSC compared to UCB ($p<0.05$; Figure 5).

Gene related to self-renewal, reprogramming and stem cell maintenance decline after transplantation with adult HSC

Differently from what observed after UCBT, a statistical significant decrease of expression of *NANOG, SOX1* and *HOXB3* ($p<0.01$) and of *OCT4, SOX2* and *HOXB4* ($p<0.05$) was observed in CD34+ cells from patients transplanted with adult HSC when compared with CD34+ cells from adult donors. There was also a trend of reduction of *DPPA2, LIN28, NES, PTEN* (Figure 5). These data seem in keeping with the loss of stemness described earlier [38].

Genes regulating self-renewal, cell reprogramming and stem cell maintenance are over-expressed in CD34+ cells after UCBT but not after adult HSCT

The over-expression of reprogramming genes (*DPPA2, LIN28, NANOG, NES, OCT4, PTEN, SOX1, SOX2, HOXB3*) shown by HSC after UCBT was not observed in HSC

from patients transplanted with adult HSC. This data remarkably emphasized how UCB CD34+ cells choose a different transcriptional asset from adult CD34+ cell when they are challenged by transplant. Although not statistically significant, the median value of *HOXB4* was higher after UCBT than after adult HSCT (Figure 5).

Some reprogramming genes are similarly expressed in CD34+ cells after UCBT and iPS although their respective signature is divergent

Because we observed that several samples of CD34+ cells after UCBT over-express genes involved in somatic cell reprogramming, we reasoned that a comparison with iPS cells was interesting and useful. The expression pattern of the reprogramming genes *DPPA2, NANOG, LIN28, OCT4, SOX2* are similarly in adult patients after UCBT and in iPS (Figure 5). To further investigate the similarities and differences between iPS and UCBT we analyzed the entire spectrum of 91 genes.

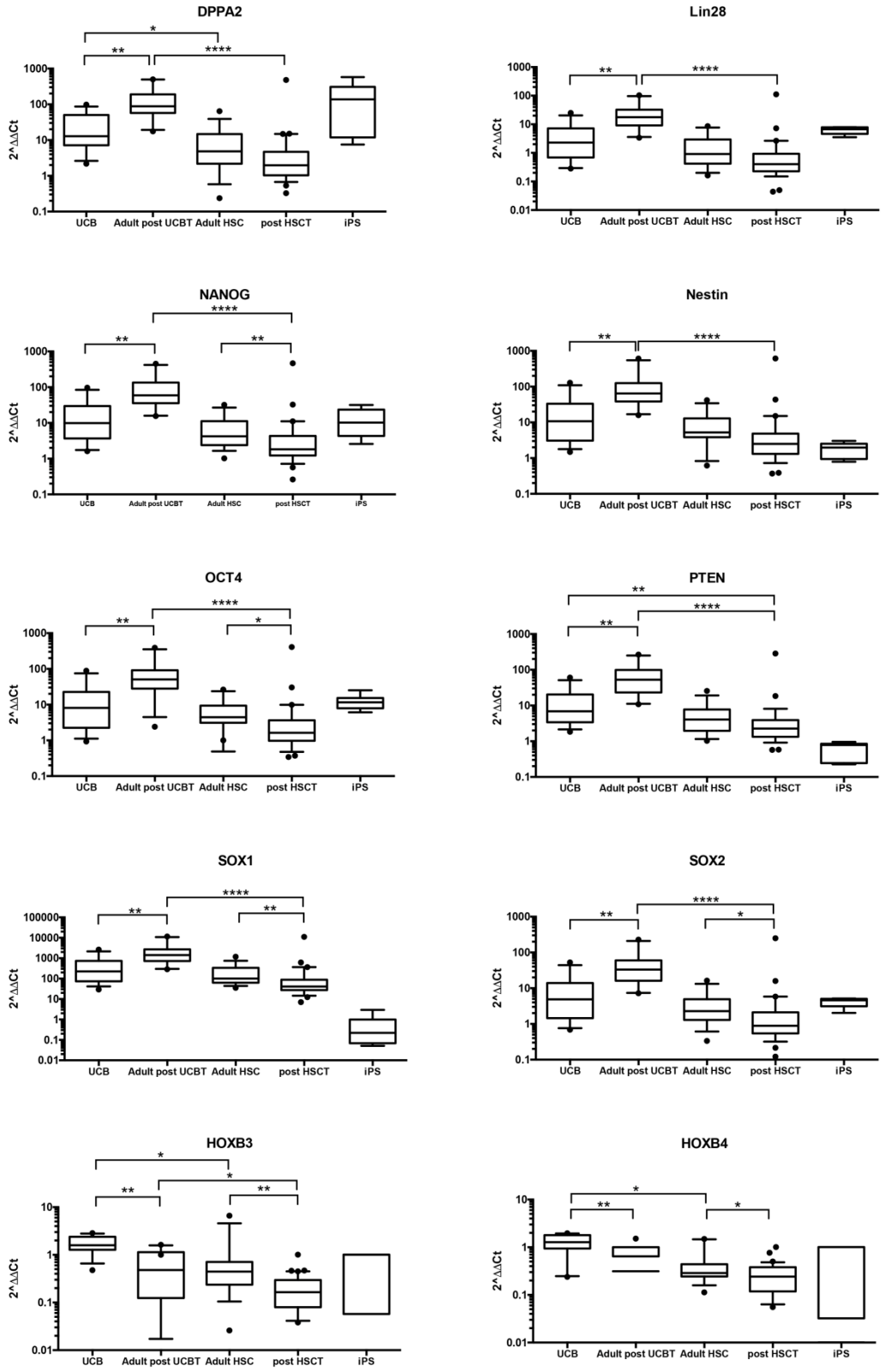


Figure 5: mRNA expression levels expressed as $2^{-\Delta\Delta Ct}$ in CD34+ cells separated from UCB units, BM cells from adult patients after UCBT, from adult healthy donors (adult HSC), from BM cells from adult and pediatric patients post HSCT and iPS. Horizontal bars indicate the median value and stars the statistical significance (* p=0,01 to 0,05; ** p=0,001 to 0,01; *** p=0,0001 to 0,001 and **** p<0,0001).

Figure 6 showed the comparison between the average expression of the genes in UCBT (*y* axis) and iPS (*x* axis). We confirmed that *DPPA2, NANOG, LIN28, OCT4, SOX2*, were expressed at similar levels as they lie close to the diagonal (*x=y*) where gene expression in UCBT cells was equal to that observed in iPS. However, there were remarkable differences in the expression pattern of many genes; for example, *PTEN SOX1* and *NES* expressions were significantly higher after UCBT cells than in iPS.

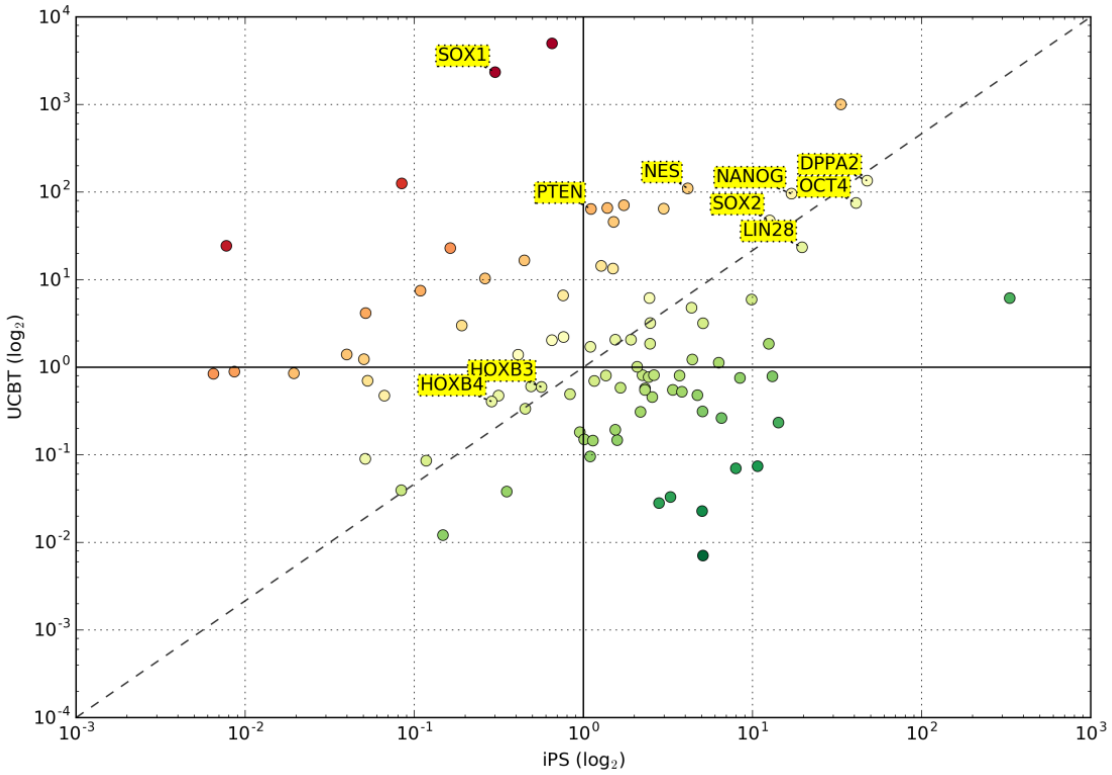


Figure 6: loglog plot comparing the average expressions of genes in UCBT (*y* axis) and iPS (*x* axis). The dashed line represents the diagonal line (*x=y*), indicating equal average expression in UCBT and iPS. DPPA2, NANOG, LIN28, OCT4, SOX2, were expressed at similar levels as they lie close to the diagonal. The expression of many other genes such as PTEN SOX1 and NES, was significantly different. The color gradient from red to green is used to indicate genes up regulated in UCBT (red) and genes up regulated in iPS.

mRNA results are confirmed by protein analysis

Immunofluorescence analysis was carried out with specific antibodies recognizing proteins coded by DPPA2, LIN28, NANOG, NES, OCT4, PTEN, SOX1, SOX2, HOXB3 and HOXB4 in 69 samples analyzed by mRNA expression and in additional 44 samples with insufficient material for gene expression analysis. Figure 7 showed that the number of cells expressing the protein was always higher in adult after UCBT. In particular, for LIN28, OCT4, NES DPPA2, SOX1 and SOX2 no positive cells were detectable in UCB, whereas for adult after UCBT the percentage of positive cells ranges from 80% to 95%. In UCB cells only for NANOG and PTEN the percentage of positive cells was being 40% and 80%, respectively. Nevertheless, the level of signal intensity in single cells was always significantly lower in UCB (see median values and ranges reported in Figure 7).

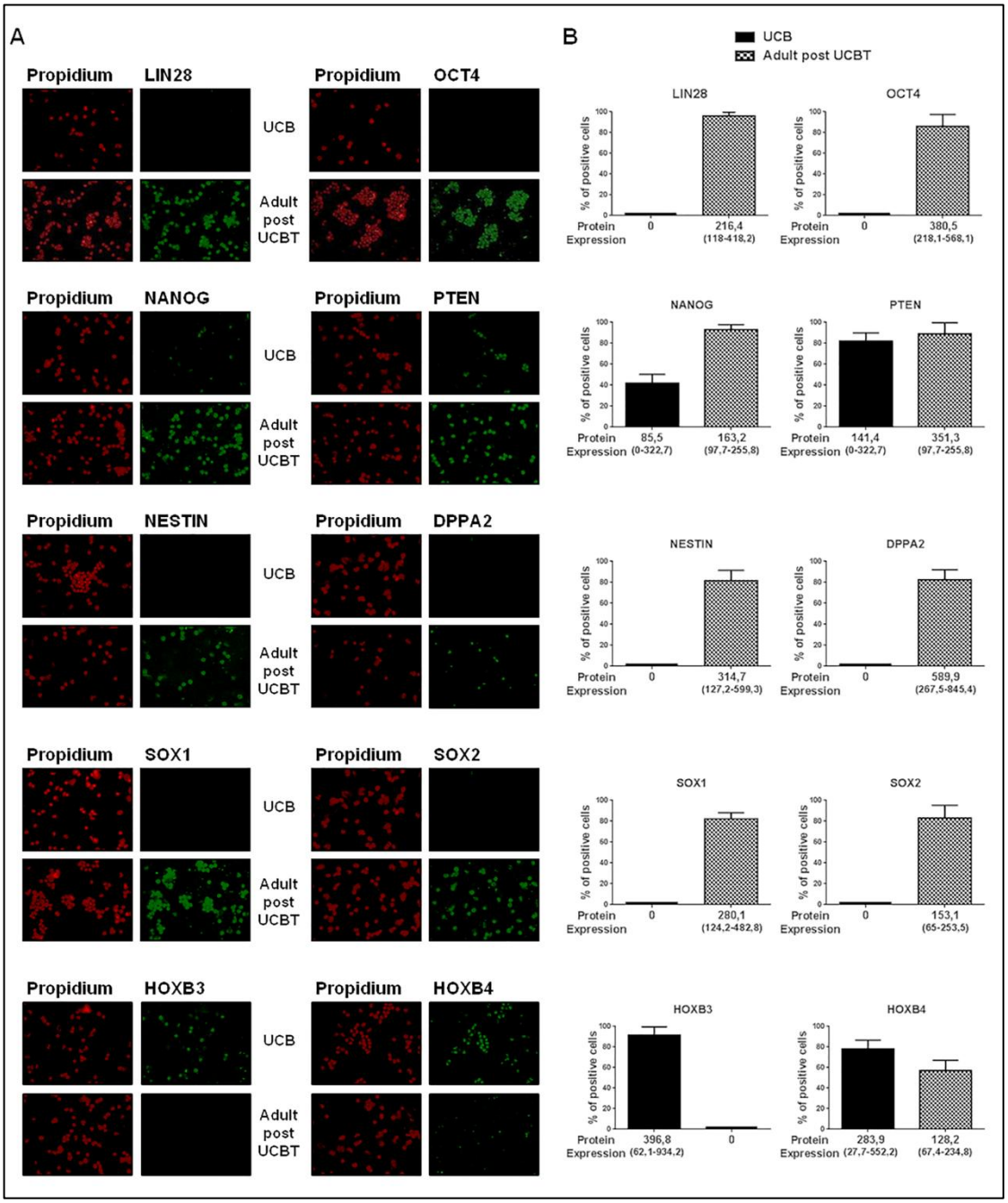


Figure 7: **Panel A:** Immunofluorescence using specific antibody (green signal) for the proteins indicated in top of the picture. Nuclei are marked in red. For each protein evaluated the upper panels are referred to CD34+ cells isolated from UCB and the lower panels are referred to CD34+ cells isolated after UCBT. **Panel B:** for each protein indicated in top of the picture the black column represents the percentage of CD34+ from UBC positive for the selected protein and the grey column the percentage of positive CD34+ from adult patients after UCBT. The mean value of protein expression in single cells and the range is indicated below the columns.

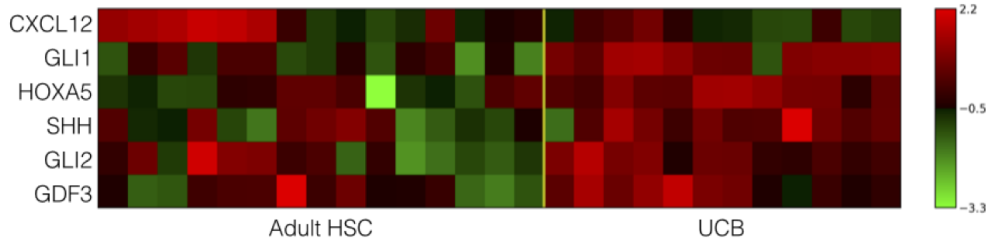
Sparse multivariate regularization analysis

To better capture the interplay among genes in the different groups (classes) of stem cell sources, we performed a sparse multivariate analysis based on l_1l_{2FS}. We compared UCB vs. adult HSC, UCB vs. adult after UCBT, adult HSC vs. adult after HSCT, and adult after UCBT vs. adult after HSCT. For each comparison, l_1l_{2FS} identifies an optimal set of weights associated to each the 91 genes. By design, l_1l_{2FS} forces some of the weights to be exactly zero, thus selecting only those genes associated with a non-null weight (gene signature). These weights are used to build a linear classifier that is associated to a prediction accuracy and an MCC score evaluating how well the selected genes are indeed able to discriminate the two classes. We recall that the accuracy was the proportion of correctly classified samples among the total number of cases examined; therefore, a good classifier has a high accuracy, possibly about 100%. The MCC score ranges between -1 and +1. The greater MCC, the better the prediction with negative score marking below chance accuracy. The signatures were presented in the form of lists of genes ranked according to a frequency score and visualized by means of a heatmap plot. (Fig. 8,10,12,14). To assess the functional association among the selected genes we presented an inferred gene network obtained with the STRING webtool, where recurring instances of neighboring genes were used to infer the associations among the genes in the signature (Fig. 9,11,13,15). Finally, to functionally characterize the gene signatures, we performed an enrichment analysis in KEGG pathway database, using the online toolkit WebGestalt, obtaining the enriched pathways (Tab. 1-4).

Only six genes are sufficient to discriminate UCB from adult HSC

To establish the global differences of UCB and adult HSC, we applied to the dataset the multivariate analysis that estimates a classifier discriminating between the two groups. The set of discriminant genes was composed of the following six genes: CXCL12, GLI1, HOXA5, SHH, GLI2 and GDF3. The associated accuracy is 89% and the MCC is 0.8. The excellent prediction performance reflects the striking inter class homogeneity of the considered samples. Figure 8 reports the selected genes (panel A) and the corresponding heatmap (panel B). Figure 9 shows the network resulting from STRING analysis. Table 1 reports the pathways we found enriched in KEGG. Among them, the Hedgehog pathway was up regulated in UCB compared to adult HSC.

A



B

Gene	Fr. (%)
CXCL12	100
GLI1	100
HOXA5	80
SHH	60
GLI2	60
GDF3	60

Figure 8 Panel A: Heatmap associated to the 6 genes of the signature discriminating between adult HSC (15 samples) and UCB (12 samples). The expression data for each individual gene have been scaled and are represented by pseudo-colors in the heatmap. Red color corresponds to high level of expression and green color corresponds to low level of expression as also shown in the colorbar. **Panel B**: Gene Signature. List of 6 gene symbols selected by the $l_1 l_{2FS}$ procedure. For each gene we reported the corresponding frequency percentage score.

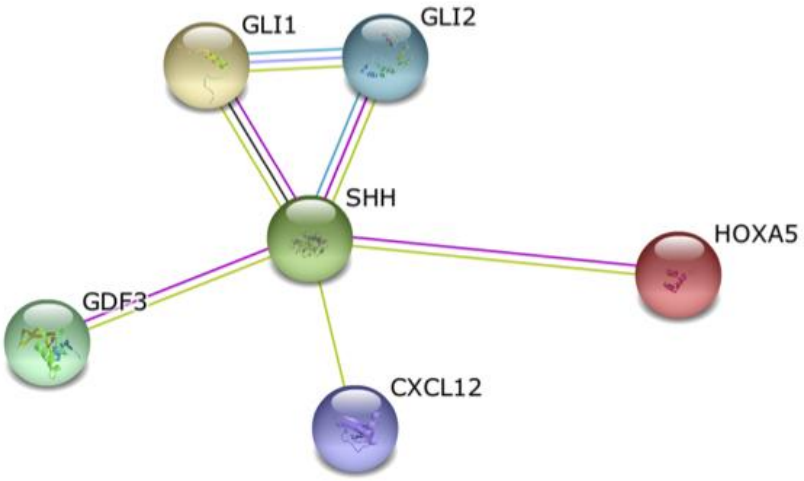


Figure 9: Association network of the genes in the signature between HSC and UCB inferred by the STRING online tool. The identified connections indicated five different data types: Textmining (lime green), Experiments (magenta), Databases (Turquoise), Coexpression (Black), Homology (purple).

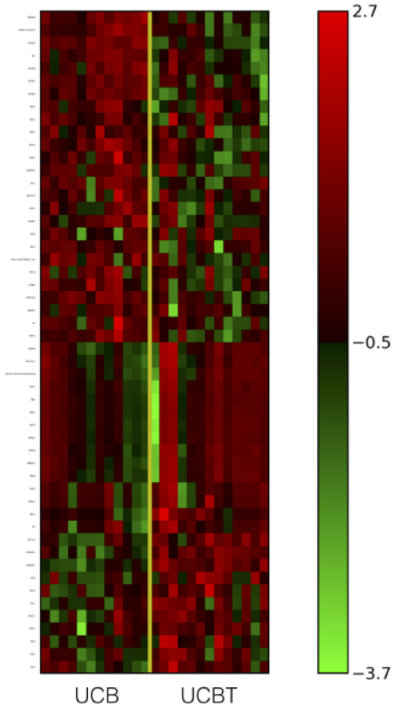
Table 1: pathways enriched in KEGG – HSC vs UCB

KEGG Pathway	#Gene	Gene Symbol
Basal cell carcinoma	**3**	GLI2, GLI1, SHH
Hedgehog signaling pathway	**3**	GLI2, GLI1, SHH
Pathways in cancer	**3**	GLI2, GLI1, SHH

A signature of fifty-two genes discriminate cord blood CD34+ cells before and after transplantation

Fifty-two genes (Figure 10 panel A), including NANOG, OCT4, PTEN, HOXB3 and HOXB4 were differentially expressed between UCB and adult after UCBT with an accuracy of 88% and an MCC value of 0.8. The corresponding heatmap is presented in Figure 10 panel B. Figure 11 presents the associated gene network inferred by STRING. Table 2 reports the pathways we founded enriched in KEGG using the identified signature. In particular, among the 52 genes the analysis extrapolated NOTCH pathway genes such as NOTCH3, NOTCH1 and HES1.

A



B

Gene	Fr. (%)	Gene	Fr. (%)
HOXA5	100	IL3	60
CDKN2A	100	NOTCH1	60
CXCL12	100	GDF3	60
KLF2	80	NOTCH3	60
CDKN2C	80	HOXB5	60
DLL3	80	BCL2L11	60
HOXB3	80	CD44	60
GLI1	60	TLE1	60
NANOG	60	CDKN1A	60
LIF	60	TCF3	60
GLI2	60	COQ7	60
HES1	60	STAT3	60
NES	60	SOX2	60
MCL1-Hs007	60	SOX1	60
LIN28	60	CXCR4	60
LIG4	60	DLL1	60
KDM5B	60	SMAD5	60
HIF1A	60	DLL4	60
KLF4	60	DPPA2	60
GFI1	60	RUNX1;SH3D19	60
KITLG	60	DPPA3	60
KIT	60	FGF4	60
XRCC5	60	PTEN	60
JAG2	60	OCT4	60
ITGB1	60	PIWIL1	60
HOXA9	60	HOXB4	60

Figure 10 Panel A: Heatmap associated to the 52 genes of the signature discriminating between UCB (12 samples) and UCBT (15 samples). The expression data for each individual gene have been scaled and was represented by pseudo-colors in the heatmap. Red color corresponds to high level of expression and green color corresponds to low level of expression as also shown in the colorbar. **Panel B**: Gene Signature. List of 52 gene symbols selected by the $l_1 l_{2FS}$ procedure. For each gene we reported the corresponding frequency percentage score.

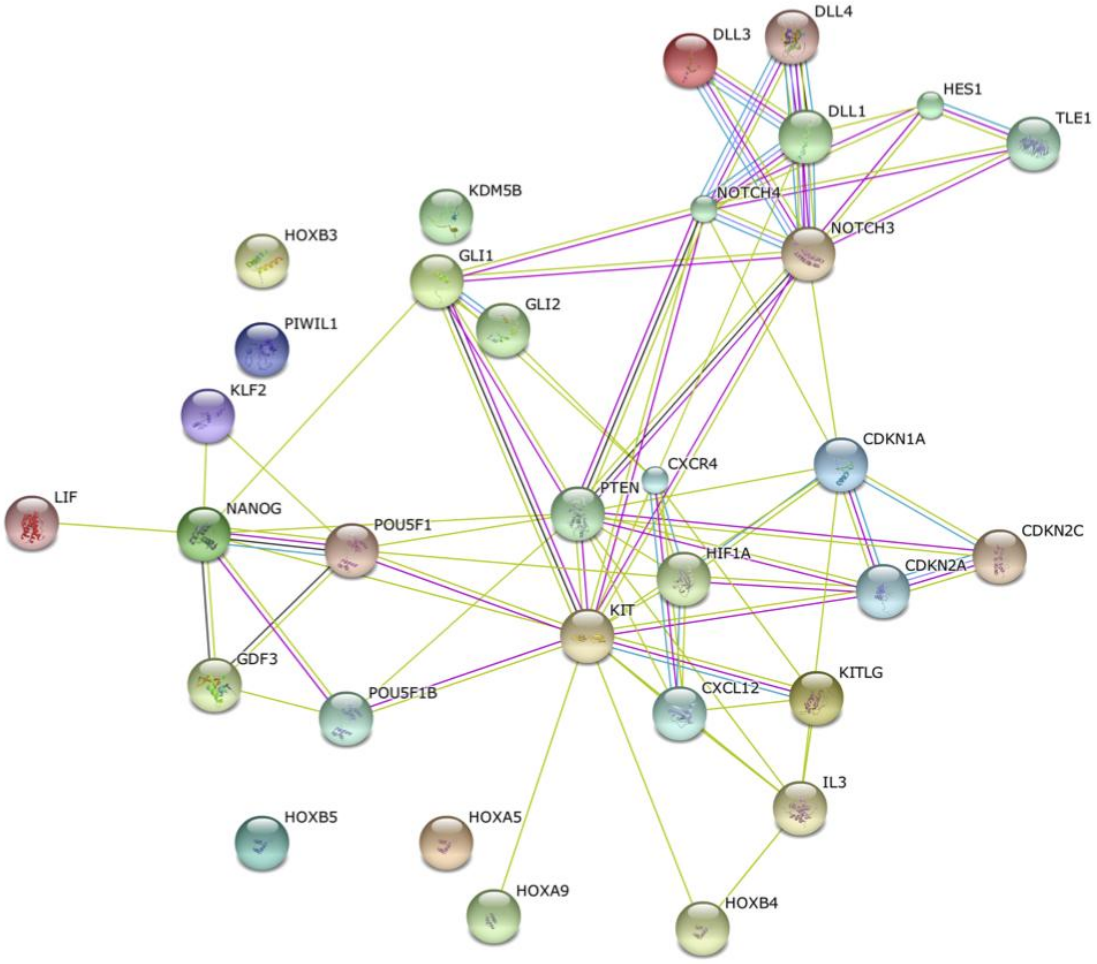


Figure 11: Association network of the genes in the signature between UCB (12 samples) and UCBT (15 samples) inferred by the STRING online tool.

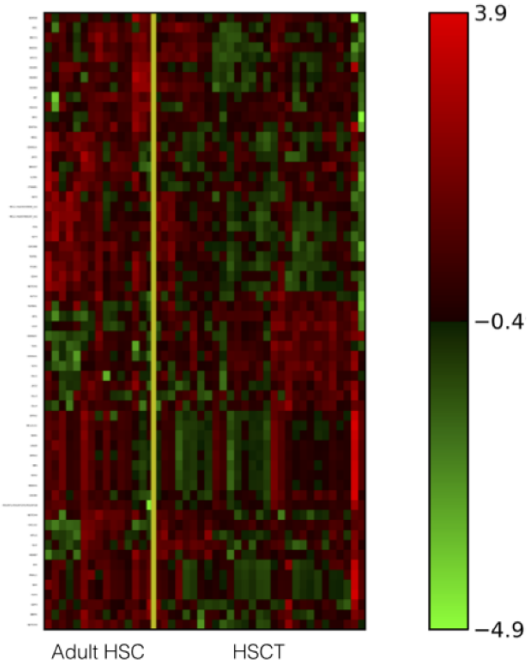
Table 2. pathways enriched in KEGG – UCB vs UCBT

KEGG Pathway	#Gene	Gene Symbol
Pathways in cancer	**12**	GLI2, HIF1A, PTEN, FGF4, ITGB1, KITLG, CDKN1A, RUNX1, KIT, STAT3, CDKN2A, GLI1,
Notch signaling pathway	**7**	HES1, NOTCH1, NOTCH3, DLL3, JAG2, DLL4, DLL1
Cytokine-cytokine receptor interaction	**6**	KITLG, LIF, IL3, CXCL12, CXCR4, KIT
Melanoma	**4**	CDKN1A, PTEN, FGF4, CDKN2A
Dorso-ventral axis formation	**3**	NOTCH1, NOTCH2, PIWIL1
Hematopoietic cell lineage	**4**	KITLG, IL3, CD44, KIT
Acute myeloid leukemia	**3**	RUNX1, KIT, STAT3
Glioma	**3**	CDKN1A, PTEN, CDKN2A
P53 signaling pathway	**3**	CDKN1A, PTEN, CDKN2A
Chronic myeloid leukemia	**3**	RUNX1, CDKN1A, CDKN2A
Leukocyte transendothelial migration	**3**	CXCL12, CXCR4, ITGB1
Axon guidance	**3**	CXCL12, CXCR4, ITGB1
Cell cycle	**3**	CDKN2C, CDN1A, CDN2A
Jak-STAT signaling pathway	**3**	LIF, IL3, STAT3
Chemokine signaling pathway	**3**	CXCL12, CXCR4, STAT3

A signature of sixty-two genes discriminate adult CD34+ cells before and after transplantation

The expression levels of 62 genes discriminate adult donors from adult and pediatric patients after transplant in with an accuracy of 88% and an MCC value of 0.7. The signature is reported in Figure 12 panel A. Interestingly, the selected genes include NES, SOX1, SOX2, OCT4, DPPA2, NANOG, LIN28, all these genes were downregulated after transplant compared to the adult HSC before transplant. The associated heatmap is shown in Figure 12 panel B. Figure 13 presents the associated gene network inferred by STRING. Table 3 reports the pathways we founded enriched in KEGG using the identified signature. The analysis extrapolated genes of the NOTCH pathway (JAG1, JAG2, NOTCH2, NOTCH3, NOTCH4, HES1) and genes of the Hedgehog pathway (GLI2, IHH, SHH, SMO).

A



B

Gene	Fr. (%)	Gene	Fr. (%)
CDKN2A	100	ERCC1	80
FOS	100	MCL1-Hs007	80
PIWIL1	100	KLF10	80
IHH	100	BCL2L11	80
SHH	100	SOX1	80
HOXB5	100	SMAD7	80
HES1	100	TGFB1	80
GFI1	100	MYC	80
HOXB3	100	RAD50	80
FGF4	100	TLE1	80
CXCL12	100	TCF3	80
ITGB1	80	SOX2	80
LIN28	80	NOTCH4	80
LIG4	80	NOTCH3	80
KLF4	80	NOTCH2	80
KLF2	80	VPS72	80
CDKN1A	80	NES	80
KIT	80	NANOG	80
CDKN2C	80	DLL4	60
BMP4	80	SMO	60
JAG1	80	STAT5A	60
IL3RA	80	DPPA2	60
GDF3	80	KDM5B	60
CSF2RB	80	HOXB7	60
DLL3	80	TGFBR1	60
HOXB4	80	DLL1	60
MCL1-Hs030	80	CXCR4	60
HOXA5	80	JAG2	60
DPPA3	80	OCT4	60
CD44	80	CTNNB1	60
GLI2	80	KITLG	60

Figure 12 Panel A: Heatmap associated to the 62 genes of the signature discriminating between adult HSC (15 samples) and adult and pediatric after HSCT (29 samples). The expression data for each individual gene have been scaled and was represented by pseudo-colors in the heatmap. Red color corresponds to high level of expression and green color corresponds to low level of expression as also shown in the colorbar. **Panel B**: Gene Signature. List of 62 gene symbols selected by the l_1l_{2FS} procedure. For each gene we reported the corresponding frequency percentage score.

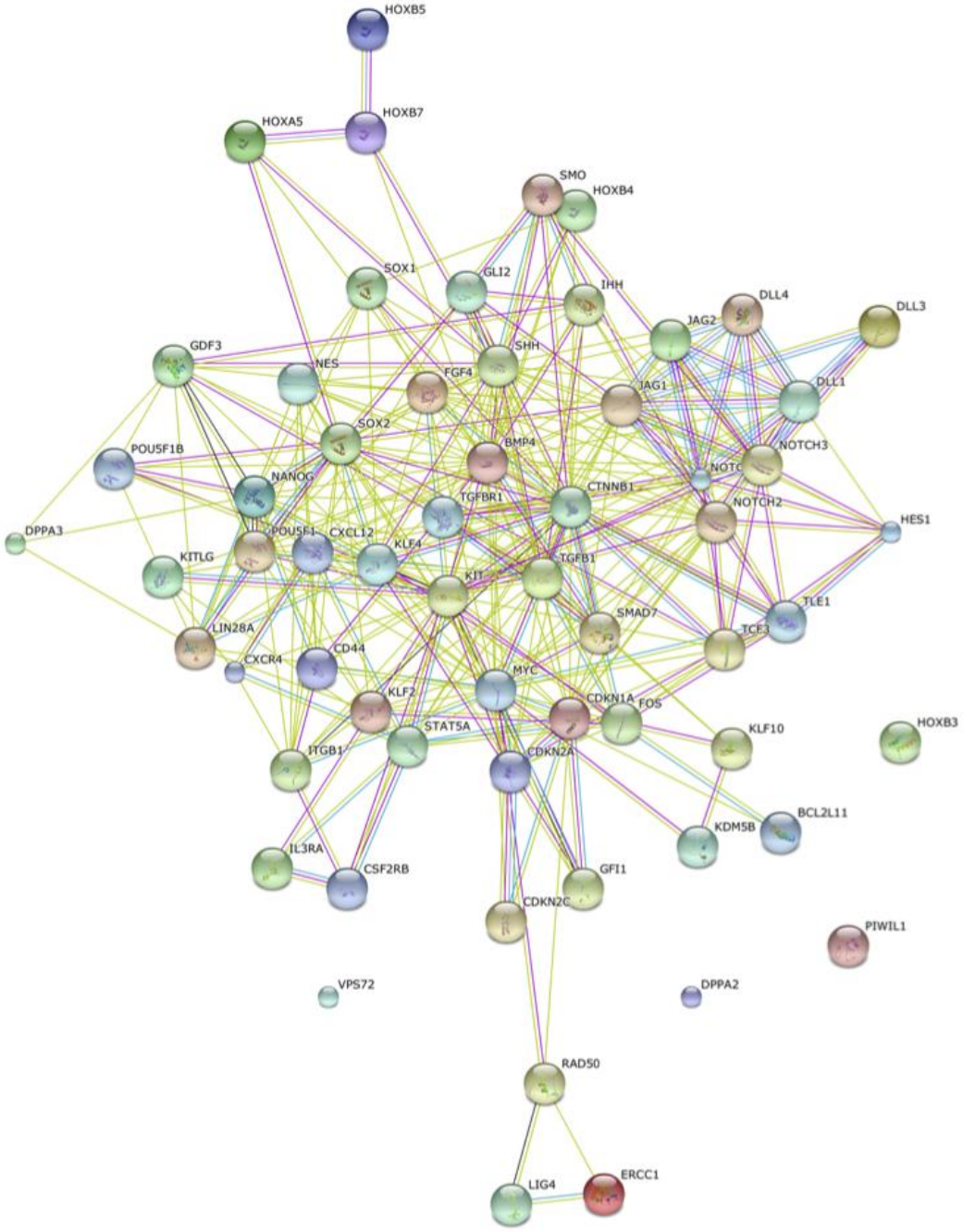


Figure 13: Association network of the genes in the signature between adult HSC (15 samples) and adult and pediatric after HSCT (29 samples) inferred by the STRING online tool.

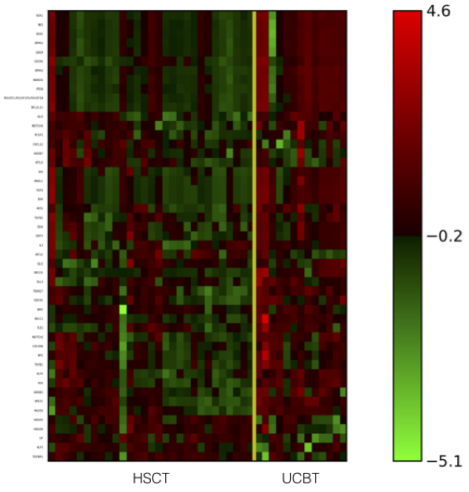
Table 3. pathways enriched in KEGG – HSC vs HSCT

KEGG Pathway	#Gene	Gene Symbol
Pathways in cancer	**16**	TGFB1, GLI2, CTNNB1, FGF4, MYC, SHH, ITGB1, SMO, KITLG, STAT5A, TGFBR1, BMP4, CDKN1A, FOS, KIT, CDKN2A
Notch signaling pathway	**9**	HES1, DLL3, JAG1, DLL4, DLL1, NOTCH4, JAG2, NOTCH2, NOTCH3
Chronic myeloid leukemia	**6**	TGFB1, STST5A, TGFBR1, CDKN1A, MYC, CDKN2A
Cytokine-cytokine receptor interaction	**8**	TGFB1, CSF2RB, CXCR4, KITLG, TGFBR1, CXCL12, KIT, IL3RA
Basal cell carcinoma	**5**	BMP4, GLI2, CTNNB1, SHH, SMO
Hedgehog signaling pathway	**5**	BMP4, GLI2, IHH, SHH, SMO
Colorectal cancer	**5**	TGFB1, TGFBR1, CTNNB1, FOS, MYC
Dorso-ventral axis formation	**4**	NOTCH3, NOTCH2, PIWIL1, NOTCH4
TGF-beta signaling pathway	**5**	TGFB1, TGFBR1, BMP4, MYC, SMAD7
Cell cycle	**5**	TGFB1, CDKN2C, CDKN1A, MYC, CDKN2A
Hematopoietic cell lineage	**4**	KILG, CD44, KIT, IL3RA
Endocytosis	**5**	TGFB1, TGFBR1, CXCR4, KIT, SMAD7
Leukocyte transendothelial migration	**4**	CXCR4, CXCL12, CTNNB1, ITGB1
Bladder cancer	**3**	CDKN1A, MYC, CDK2A
MAPK signaling pathway	**5**	TGFB1, TGFBR1, FOS, FGF4, MYC
Intestinal immune network for IgA production	**3**	TGFB1, CXCR4, CXCL12
Jak-STAT signaling pathway	**4**	CSF2RB, STAT5A, IL3RA, MYC
Acute myeloid leukemia	**3**	STAT5A, KIT, MYC
Melanoma	**3**	CDKN1A, FGF4, CDKN2A
Leishmaniasis	**3**	TGFB1, FOS, ITGB1
Pancreatic cancer	**3**	TGFB1, TGFBR1, CDKN2A
ErbB signaling pathway	**3**	STAT5A, CDKN1A, MYC
Rheumatoid arthritis	**3**	TGFB1, CXCL12, FOS
Chagas disease	**3**	TGFB1, TGFBR1, FOS
Melanogenesis	**3**	KITLG, CTNNB1, KIT
Osteoclast differentiation	**3**	TGFB1, TGFBR1, FOS
Axon guidance	**3**	CXCR4, CXCL12, ITGB1

After transplantation adult versus cord blood CD34+ cells acquire a divergent transcriptional asset; reprogramming genes are a relevant part of such difference.

Finally, the comparison between UCBT and CD34+ after HSCT identified a signature of 49 genes including NES, NANOG, LIN28, OCT4, DPPA2, SOX1, SOX2 and PTEN which is shown in Figure 14 panel A together with its heatmap (panel B). This genes list distinguishes patients according to the type of stem cells they received with an accuracy of 85% and an MCC value of 0.7. Table 4 reports the pathways we founded enriched in KEGG using the identified signature. Among the identified genes the analysis extrapolated genes of the NOTCH pathway (HES1, DLL3, NOTCH2, NOTCH4) and of the Hedgehog pathway (SHH, IHH, DHH, GLI1, GLI2, SMO).

A



B

Gene	Fr. (%)	Gene	Fr. (%)
XRCC6	100	SHH	100
DHH	100	VPS72	100
NOTCH4	100	TRIM27	100
NES	100	TLE1	100
NANOG	100	BCL2L11	100
DPPA3	100	SOX1	100
LIN28	100	SOX2	100
CXCL12	100	RAD50	80
LIF	100	HIF1A	80
KLF4	100	HOXA9	80
CXCR4	100	SMO	80
IL3	100	KLF2	80
OCT4	100	HOXB3	80
IHH	100	NOTCH2	80
DLL3	100	HOXB7	80
HOXA5	100	SPI1	60
HES1	100	TGFB2	60
GLI2	100	GSK3A	60
GDF3	100	TGFB1	60
FOS	100	TGFBR1	60
DPPA2	100	GLI1	60
FGF4	100	KITLG	60
PIWIL1	100	PCGF2	60
ERCC1	100	CSF2RB	60
PTEN	100		

Figure 14 Panel A: Heatmap associated to the 49 genes of the signature discriminating between adult after UCBT (13 samples) and adult and pediatric after HSCT (29 samples). The expression data for each individual gene have been scaled and was represented by pseudo-colors in the heatmap. Red color corresponds to high level of expression and green color corresponds to low level of expression as also shown in the color bar. **Panel B**: Gene Signature. List of 49 gene symbols selected by the $l_1 l_{2FS}$ procedure. For each Gene we reported the corresponding frequency percentage score.

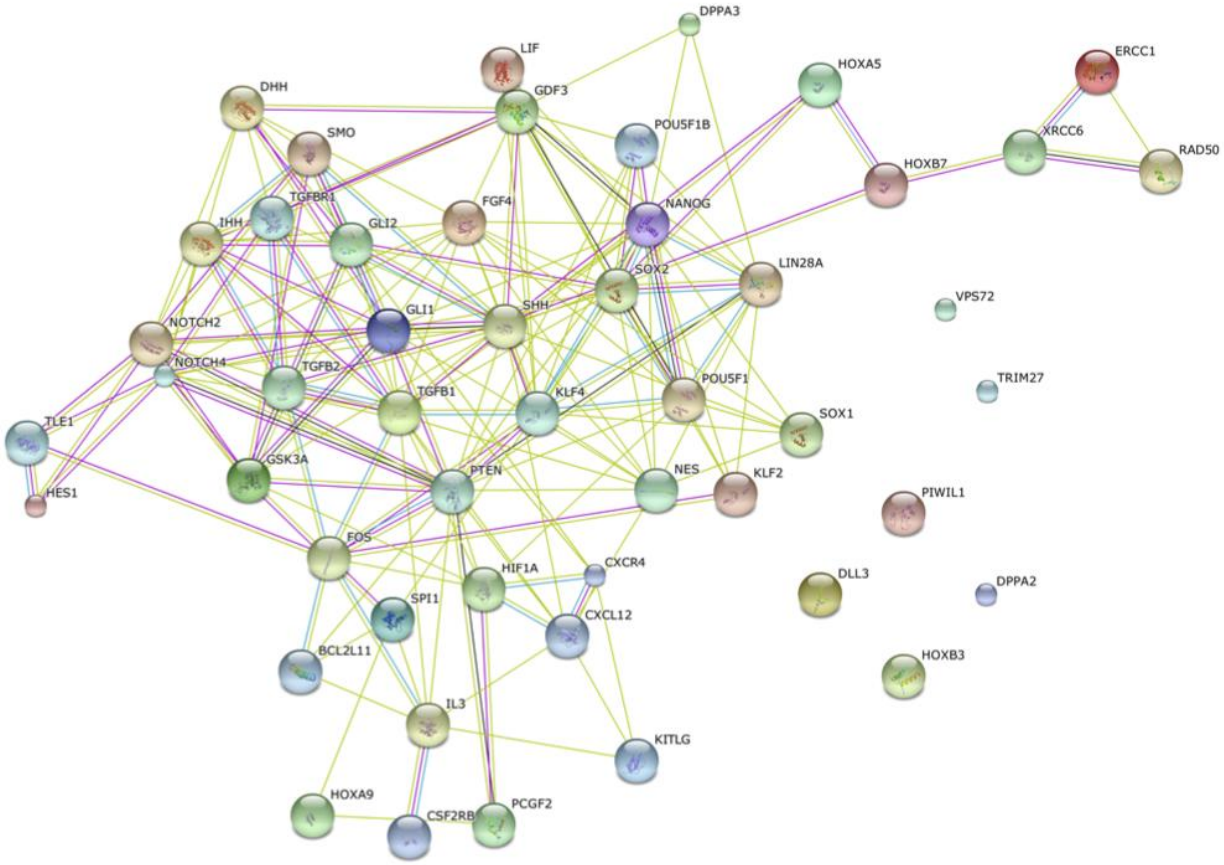


Figure 15: Association network of the genes in the signature between adult after UCBT (13 samples) and adult and pediatric after HSCT (29 samples) inferred by the STRING online tool.

Table 4. pathways enriched in KEGG – UCBT vs HSCT

KEGG Pathway	#Gene	Gene Symbol
Pathways in cancer	**13**	**TGFB1, GLI2, PTEN, HIF1A, FGF4, SHH, SMO, KITLG,TGFBR1, TGFB2, SPI1, FOS, GLI1**
Cytokine-cytokine receptor interaction	**9**	**CSF2RB, LIF, IL3, TGFB1, CXCR4, KITLG, TGFBR1, CXCL12, TGFB2**
Hedgehog signaling pathway	**6**	**DHH, GLI2, GLI1, IHH, SHH, SMO**
Notch signaling pathway	**4**	**HES1, DLL3, NOTCH2, NOTCH4**
Osteoclast differentiation	**5**	**TGFB1, TGFBR1, TGFB2, SPI1, FOS**
Basal cell carcinoma	**4**	**GLI2, GLI1, SHH, SMO**
Colorectal cancer	**4**	**TGFB1, TGFBR1, TGFB2, FOS**
Dorso-ventral axis formation	**3**	**NOTCH2, PIWIL1, NOTCH4**
Rheumatoid arthritis	**4**	**TGFB1, CXCL12, TGFB2, FOS**
Chagas disease	**4**	**TGFB1, TGFBR1, TGFB2, FOS**
MAPK signaling pathway	**5**	**TGFB1, TGFBR1, TGFB2, FOS, FGF4**
Intestinal immune network for IgA production	**3**	**TGFB1, CXCL12, CXCR4**
Renal cell carcinoma	**3**	**TGFB1, TGFB2, HIF1A**
Pancreatic cancer	**3**	**TGFB1, TGFBR1, TGFB2**
Chronic myeloid leukemia	**3**	**TGFB1, TGFBR1, TGFB2**
Leishmaniasis	**3**	**TGFB1, TGFB2, FOS**
Endocytosis	**4**	**TGFB1, TGFBR1, CXCR4, TGFB2**
TGF-beta signaling pathway	**3**	**TGFB1, TGFBR1, TGFB2**
Jak-STAT signaling pathway	**3**	**CSF2RB, LIF, IL3**
Chemokine signaling pathway	**3**	**CXCL12, CXCR4, GSK3A**

CD34+ cells from UCB *ex vivo* expanded show a similar behavior to Adult post UCBT

To investigate the role of the greater expansion needed to UCB stem cells after transplant, we evaluated the mRNA expression of *DPPA2, LIN28, NANOG, NES, OCT4, PTEN, SOX1 and SOX2* comparing "UCB stem cells" with "UCB stem cells *ex vivo* expanded", with "Adult post UCBT" and with "Adult transplanted with *ex vivo* expanded UCB". Figure 16 shows that "UCB stem cells *ex vivo* expanded" over-express these genes, except for *PTEN*, similar to "Adult post UCBT". Furthermore, "Adult transplanted with stem cells *ex vivo* expanded" showed expression levels comparable with "UCB ex vivo expanded". These results suggested that the UCB CD34+ cells reactivate stemness pathways to mediated expansion mandatory to repopulate an adult BM.

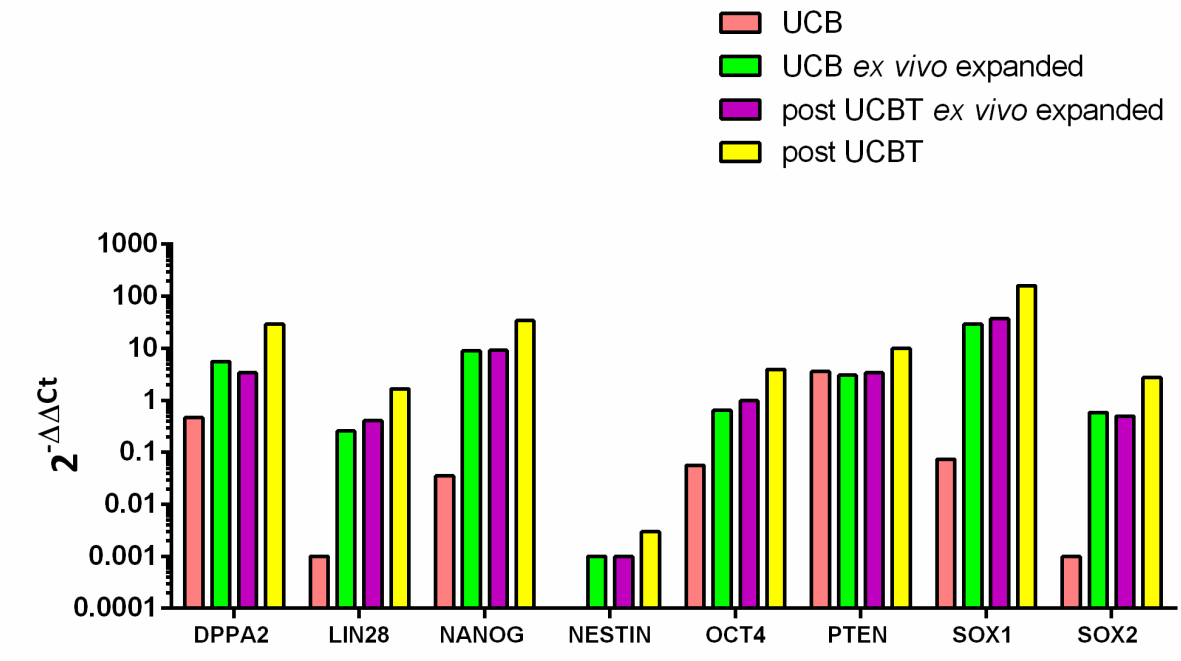


Figure 16: mRNA expression levels expressed as $2^{-\Delta\Delta Ct}$ in CD34+ cells separated from UCB units (pink), UCB *ex vivo* expanded (green), Adult post UCBT *ex vivo* expanded (purple) and Adult post UCBT (yellow).

Infusion procedure could affect stem genes expression

To clarify the possible role of infusion procedure, we also evaluated the mRNA expression of *DPPA2, LIN28, NANOG, NES, OCT4, PTEN, SOX1 and SOX2* in CD34+ cells from patients transplanted with UCB intra-bone (IB-UCBT) or endo-venous (EV). The results showed that only patients transplanted with IB-UCBT over-express stemness genes compared with UCB (Fig. 17), suggesting a role of infusion procedure.

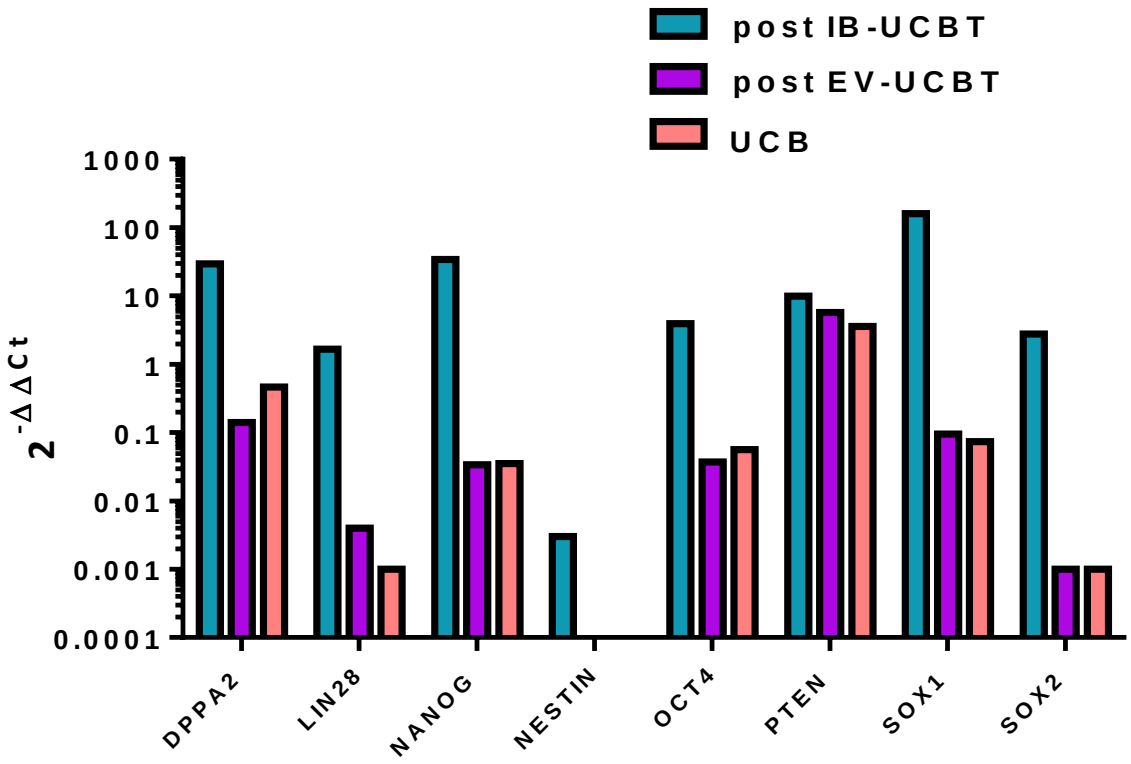


Figure 17: mRNA expression levels expressed as $2^{-\Delta\Delta Ct}$ in CD34+ cells separated from Adult post IB-UCBT (blue), Adult post EV-UCBT (purple) and UCB (pink).

DISCUSSION

The HSC ability is that of ensuring and maintaining normal peripheral blood counts; this peculiarity is also preserved when HSC are placed in a "pancytopenic" patient. Clinical observations suggest that, after a successful transplantation, hematopoiesis is fully reconstituted and completely similar to that of normal healthy donors [60]. The whole system tightly regulated as witnessed by the stabilization of CD34+ cell frequency in the bone marrow. However, after more than a million of hematopoietic transplants [60], a question is missing: are transplanted cord blood cells becoming like bone marrow cells? Namely, do UCB HSCs undergo acceleration in aging and become BM HSC few months after transplantation?

The results of our study contribute to give a partial answer to those questions and document that after transplant, HSC undergo a profound change in their transcriptional asset while, functionally, remaining adherent to the commitment to hematopoietic lineage. Previous studies aimed to explore the function of single genes involved in self-renewal utilized constitutional over-expression, knock-out or silencing of selected genes. An elegant example of such experimental approach was to enhance *in-vivo* the HSC proliferative potential by generating HOXB4-Transduced Hematopoietic Stem Cells [61]. An increased stem cell pool was obtained in mice transplanted with transduced cells. This was evident only when HSC were challenged in further transplants, whereas HOXB4-transduced HSC did not expand above levels normally observed in un-manipulated mice. This data indicate that its over-expression does not override the regulatory mechanisms that maintain the HSC pool size within normal limits. Our approach discloses a photography (snap-shot) of how HSC orchestrates their transcriptional asset to reconstitute the hematopoietic system. Our results identified DPPA2, NANOG,

PTEN, OCT4, NES, SOX1, SOX2, LIN28 as genes significantly over-expressed in several samples of CD34+ cells obtained from patients transplanted with UCB cells. Among these genes, OCT4, SOX2, NANOG and LIN28 are the master regulators of pluripotency in embryonic stem cells (ESC). They are proved critical for ESC maintenance and capable of reprogramming mature somatic cells and to induce iPS [62-66]. To our knowledge, this behavior is unique of transplanted UCB cells and does not occur after adult HSC transplantation and suggest a reactivation of stemness pathway (ability) in UCB cells transplanted. Moreover, we noticed a remarkable variability in the expression of reprogramming genes among samples and we cannot provide a ready explanation for this phenomenon.

We identified several genes differentially expressed with respect to the grafted population of CD34+ cells. Immunofluorescence assay confirmed the results of mRNA over-expression by showing high levels of the corresponding proteins. By measuring proteins at single cell level, we found that the majority of BM CD34+ cells after UCBT shows a clear over-expression of the proteins involved in self-renewal and re-programming although with a high degree of variability from cell to cell thus reflecting the heterogeneity of CD34+ cell population. In contrast, in native UCB cells the level of these protein is too low to be detectable by this method strengthening our observation on the remodulation of these protein in UCBT; As consequence of HOX gene experiments referred above [61], we would have expected no difference in genes expression dependent from hematopoietic cell source. Unexpectedly, in our analysis, many genes recruited by CD34+ cells after UCBT are different when compared to transplant with adult cells. The robustness and consistency of this data is provided by multivariate learning algorithm showing that each cell source analyzed has a very specific signature.

Each signature was characterized by an excellent prediction capability and by an accuracy ranging from 80 to 90%. When we found the over-expression of genes and proteins involved in reprogramming, we did not have a reference level of expression associated with specific functional properties; thus, we investigated gene expression in different iPS cell lines, that are normally created expressing these genes exogenically. In CD34+ cells after UCBT, genes considered crucial for re-programming (NANOG, OCT4, SOX2, LIN28) showed a pattern of expression nearly super imposable in iPS cells. The acquisition of such a signature by CD34+ cells after UCBT is a remarkable finding and represents, so far, a unique example of "spontaneous" expression of reprogramming genes by somatic cells in adult life. In this context, it might be interesting to refer to the high propensity of UCB cells to be reprogrammed to iPS by the transduction of only two factors such as OCT4 and SOX2 [67]. In the light of our findings, the slower recovery of blood values observed after UCBT could be interpreted as a resistance to differentiation and maturation of UCB HSC. Whatever the case, as clearly showed by the regression analysis (Figure 6) comparing gene expression from HSC after UCBT and iPS, most genes are differentially expressed in the two cell types. This data, while excluding that CD34+ cells after UCBT have become like iPS, suggest that upon UCBT CD34+ cells acquire a transcriptional asset that might render them more prone to plasticity. However, in spite of this potentiality HSC remain strongly anchored to hematopoietic lineage commitment. It is possible that after transplant the majority of CD34+ cells activate self-renewal and pluripotency genes to cope with expansion but fail to down-regulate somatic and lineage specification genes, making these cells refractory to re-programming. There may also be unknown mechanisms in the bone marrow microenvironment that prevent down regulation

of lineage commitment gene expression program [66]. At the same time, the transcriptional changes that HSC undergo during transplantation may be informative to understand how HSC govern their gene machinery in such a forced proliferation avoiding the risk of transformation. PTEN expression, for example, is lower in iPS as compared to CD34+ after UCBT. PTEN was reported not to be required for HSC self-renewal in fetal hematopoietic stem cells, however it is critical in adult HSC maintenance [43, 68, 69]. More intriguingly, its loss of function in adult HSC can promote leukemogenesis over time [70] and represents one of the major obstacles in utilizing iPS cells (or ES cells) for clinical purposes. We found that PTEN was significantly up-regulated after UCBT compared to iPS. Since it is clearly evident that CD34+ cells after UCBT do not generate tumors, the present study may represent an initial platform to learn how CD34+ cells are orchestrating their gene expression by combining expansion and, at the same time, avoiding transformation. Since UCB CD34+ cells are more prone to activate reprogramming genes than adult HSC, it is possible that they utilized this machinery to cope with the challenge of such tremendous expansion. Perhaps, by learning how transplanted cells behave during hematopoietic regeneration we could gain useful information in designing new methods to expand HSC ex-vivo.
This study places the question of understanding why to achieve hematopoietic reconstitution UCB HSC choose a different transcriptional asset from adult HSC including the recruitment of reprogramming genes and, at the same time, remain adherent to hematopoietic lineage. These results reveal undisclosed characteristics and potentialities of hematopoietic stem cells and open a new area of research in transplantation biology.

REFERENCES

1. Jagannathan-Bogdan M and Zon LI. Hematopoiesis. Development. 2013; 140(12):2463-2467.
2. Birbrair A and Frenette PS. Niche heterogeneity in the bone marrow. Annals of the New York Academy of Sciences. 2016; 1370(1):82-96.
3. Bornhauser M, Illmer T, Schaich M, Soucek S, Ehninger G, Thiede C and group ASs. Improved outcome after stem-cell transplantation in FLT3/ITD-positive AML. Blood. 2007; 109(5):2264-2265; author reply 2265.
4. Perrimon N, Pitsouli C and Shilo BZ. Signaling mechanisms controlling cell fate and embryonic patterning. Cold Spring Harbor perspectives in biology. 2012; 4(8):a005975.
5. Guruharsha KG, Kankel MW and Artavanis-Tsakonas S. The Notch signalling system: recent insights into the complexity of a conserved pathway. Nature reviews Genetics. 2012; 13(9):654-666.
6. Delaney C, Heimfeld S, Brashem-Stein C, Voorhies H, Manger RL and Bernstein ID. Notch-mediated expansion of human cord blood progenitor cells capable of rapid myeloid reconstitution. Nature medicine. 2010; 16(2):232-236.
7. Calvi LM, Adams GB, Weibrecht KW, Weber JM, Olson DP, Knight MC, Martin RP, Schipani E, Divieti P, Bringhurst FR, Milner LA, Kronenberg HM and Scadden DT. Osteoblastic cells regulate the haematopoietic stem cell niche. Nature. 2003; 425(6960):841-846.
8. Zhang J, Niu C, Ye L, Huang H, He X, Tong WG, Ross J, Haug J, Johnson T, Feng JQ, Harris S, Wiedemann LM, Mishina Y and Li L. Identification of the haematopoietic stem cell niche and control of the niche size. Nature. 2003; 425(6960):836-841.
9. Ding L and Morrison SJ. Haematopoietic stem cells and early lymphoid progenitors occupy distinct bone marrow niches. Nature. 2013; 495(7440):231-235.
10. Rieger MA and Schroeder T. Hematopoiesis. Cold Spring Harbor perspectives in biology. 2012; 4(12).
11. Holowiecki J. Indications for hematopoietic stem cell transplantation. Polskie Archiwum Medycyny Wewnetrznej. 2008; 118(11):658-663.
12. Petersdorf EW, Gooley TA, Anasetti C, Martin PJ, Smith AG, Mickelson EM, Woolfrey AE and Hansen JA. Optimizing outcome after unrelated marrow transplantation by comprehensive matching of HLA class I and II alleles in the donor and recipient. Blood. 1998; 92(10):3515-3520.
13. Hansen JA, Gooley TA, Martin PJ, Appelbaum F, Chauncey TR, Clift RA, Petersdorf EW, Radich J, Sanders JE, Storb RF, Sullivan KM and Anasetti C. Bone marrow transplants from unrelated donors for patients with chronic myeloid leukemia. The New England journal of medicine. 1998; 338(14):962-968.
14. Aversa F, Reisner Y and Martelli MF. Hematopoietic stem cell transplantation from alternative sources in adults with high-risk acute leukemia. Blood cells, molecules & diseases. 2004; 33(3):294-302.
15. Rocha V, Labopin M, Gluckman E, Powles R, Arcese W, Bacigalupo A, Reiffers J, Iriondo A, Ringden O, Ruutu T, Frassoni F, Acute Leukemia Working Party of the European B and Marrow Transplant R. Relevance of bone marrow cell dose on

allogeneic transplantation outcomes for patients with acute myeloid leukemia in first complete remission: results of a European survey. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2002; 20(21):4324-4330.
16. Anguita-Compagnon AT, Dibarrart MT, Palma J, Paredes L, Mosso C, Montalva R, Salas L, Araos D, Delgado I and Majlis A. Mobilization and collection of peripheral blood stem cells: guidelines for blood volume to process, based on CD34-positive blood cell count in adults and children. Transplantation proceedings. 2010; 42(1):339-344.
17. Cesaro S, Tintori V, Nesi F, Schiavello E, Calore E, Dallorso S, Migliavacca M, Capolsini I, Desantis R, Caselli D, Fagioli F, Luksch R, Panizzolo I, Tridello G and Prete A. A prospective study on the efficacy of mobilization of autologous peripheral stem cells in pediatric oncohematology patients. Transfusion. 2013; 53(7):1501-1509.
18. Lauber S, Latta M, Kluter H and Muller-Steinhardt M. The Mannheim Cord Blood Bank: Experiences and Perspectives for the Future. Transfusion medicine and hemotherapy : offizielles Organ der Deutschen Gesellschaft fur Transfusionsmedizin und Immunhamatologie. 2010; 37(2):90-97.
19. Armson BA, Maternal/Fetal Medicine Committee SoO and Gynaecologists of C. Umbilical cord blood banking: implications for perinatal care providers. Journal of obstetrics and gynaecology Canada : JOGC = Journal d'obstetrique et gynecologie du Canada : JOGC. 2005; 27(3):263-290.
20. Reiberger T, Payer BA, Ferlitsch A, Sieghart W, Breitenecker F, Aichelburg MC, Schmied B, Rieger A, Trauner M, Peck-Radosavljevic M, Vienna Hepatic Hemodynamic L, Vienna HIV and Liver Study G. A prospective evaluation of pulmonary, systemic and hepatic haemodynamics in HIV-HCV-coinfected patients before and after antiviral therapy with pegylated interferon and ribavirin. Antiviral therapy. 2012; 17(7):1327-1334.
21. Schonberger S, Niehues T, Meisel R, Bernbeck B, Laws HJ, Kogler G, Enzmann J, Wernet P, Gobel U and Dilloo D. Transplantation of haematopoietic stem cells derived from cord blood, bone marrow or peripheral blood: a single centre matched-pair analysis in a heterogeneous risk population. Klinische Padiatrie. 2004; 216(6):356-363.
22. Petersdorf EW, Longton GM, Anasetti C, Martin PJ, Mickelson EM, Smith AG and Hansen JA. The significance of HLA-DRB1 matching on clinical outcome after HLA-A, B, DR identical unrelated donor marrow transplantation. Blood. 1995; 86(4):1606-1613.
23. Waller-Wise R. Umbilical cord blood: information for childbirth educators. The Journal of perinatal education. 2011; 20(1):54-60.
24. Moise KJ, Jr. Umbilical cord stem cells. Obstetrics and gynecology. 2005; 106(6):1393-1407.
25. Drew D. Umbilical cord blood banking. Advance for nurse practitioners. 2005; 13(4):Suppl 2-6; quiz 7.
26. Sullivan MJ. Banking on cord blood stem cells. Nature reviews Cancer. 2008; 8(7):555-563.
27. Hess S. Cord blood banking: a misunderstood option. Advance for nurse practitioners. 1997; 5(8):80.
28. Gunning J. Umbilical cord cell banking: an issue of self-interest versus altruism. Medicine and law. 2007; 26(4):769-780.

29. Ballen K, Broxmeyer HE, McCullough J, Piaciabello W, Rebulla P, Verfaillie CM and Wagner JE. Current status of cord blood banking and transplantation in the United States and Europe. Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation. 2001; 7(12):635-645.
30. Bornstein R, Flores AI, Montalban MA, del Rey MJ, de la Serna J and Gilsanz F. A modified cord blood collection method achieves sufficient cell levels for transplantation in most adult patients. Stem cells. 2005; 23(3):324-334.
31. Hofmeister CC, Zhang J, Knight KL, Le P and Stiff PJ. Ex vivo expansion of umbilical cord blood stem cells for transplantation: growing knowledge from the hematopoietic niche. Bone marrow transplantation. 2007; 39(1):11-23.
32. Zhai QL, Qiu LG, Li Q, Meng HX, Han JL, Herzig RH and Han ZC. Short-term ex vivo expansion sustains the homing-related properties of umbilical cord blood hematopoietic stem and progenitor cells. Haematologica. 2004; 89(3):265-273.
33. Iscove NN and Nawa K. Hematopoietic stem cells expand during serial transplantation in vivo without apparent exhaustion. Current biology : CB. 1997; 7(10):805-808.
34. Harrison DE. Proliferative capacity of erythropoietic stem cell lines and aging: an overview. Mechanisms of ageing and development. 1979; 9(5-6):409-426.
35. Harrison DE, Stone M and Astle CM. Effects of transplantation on the primitive immunohematopoietic stem cell. The Journal of experimental medicine. 1990; 172(2):431-437.
36. Siminovitch L, Till JE and McCulloch EA. Decline in Colony-Forming Ability of Marrow Cells Subjected to Serial Transplantation into Irradiated Mice. Journal of cellular and comparative physiology. 1964; 64:23-31.
37. Mauch P, Rosenblatt M and Hellman S. Permanent loss in stem cell self renewal capacity following stress to the marrow. Blood. 1988; 72(4):1193-1196.
38. Harrison DE, Astle CM and Delaittre JA. Loss of proliferative capacity in immunohemopoietic stem cells caused by serial transplantation rather than aging. The Journal of experimental medicine. 1978; 147(5):1526-1531.
39. Bell DR and Van Zant G. Stem cells, aging, and cancer: inevitabilities and outcomes. Oncogene. 2004; 23(43):7290-7296.
40. Frassoni F, Podesta M, Maccario R, Giorgiani G, Rossi G, Zecca M, Bacigalupo A, Piaggio G and Locatelli F. Cord blood transplantation provides better reconstitution of hematopoietic reservoir compared with bone marrow transplantation. Blood. 2003; 102(3):1138-1141.
41. Park IH, Zhao R, West JA, Yabuuchi A, Huo H, Ince TA, Lerou PH, Lensch MW and Daley GQ. Reprogramming of human somatic cells to pluripotency with defined factors. Nature. 2008; 451(7175):141-146.
42. Warlich E, Kuehle J, Cantz T, Brugman MH, Maetzig T, Galla M, Filipczyk AA, Halle S, Klump H, Scholer HR, Baum C, Schroeder T and Schambach A. Lentiviral vector design and imaging approaches to visualize the early stages of cellular reprogramming. Molecular therapy : the journal of the American Society of Gene Therapy. 2011; 19(4):782-789.
43. Zhang J, Grindley JC, Yin T, Jayasinghe S, He XC, Ross JT, Haug JS, Rupp D, Porter-Westpfahl KS, Wiedemann LM, Wu H and Li L. PTEN maintains haematopoietic stem cells and acts in lineage choice and leukaemia prevention. Nature. 2006; 441(7092):518-522.

44. Cilloni D, Carturan S, Bracco E, Campia V, Rosso V, Torti D, Calabrese C, Gaidano V, Niparuck P, Favole A, Signorino E, Iacobucci I, Morano A, De Luca L, Musto P, Frassoni F, et al. Aberrant activation of ROS1 represents a new molecular defect in chronic myelomonocytic leukemia. Leukemia research. 2013; 37(5):520-530.
45. Zou HaH, T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society. 2005.
46. De Mol C MS, Traskine M, Verri A. A regularized method for selecting nested groups of relevant genes from microarray data. Journal of computational biology. 2009.
47. De Mol C DVE, Rosasco L. Elastic net regularization in learning theory. Journal of Complexity. 2009.
48. Mascelli S, Barla A, Raso A, Mosci S, Nozza P, Biassoni R, Morana G, Huber M, Mircean C, Fasulo D, Noy K, Wittemberg G, Pignatelli S, Piatelli G, Cama A, Garre ML, et al. Molecular fingerprinting reflects different histotypes and brain region in low grade gliomas. BMC cancer. 2013; 13:387.
49. Squillario M and Barla A. A computational procedure for functional characterization of potential marker genes from molecular data: Alzheimer's as a case study. BMC medical genomics. 2011; 4:55.
50. Fardin P, Cornero A, Barla A, Mosci S, Acquaviva M, Rosasco L, Gambini C, Verri A and Varesio L. Identification of multiple hypoxia signatures in neuroblastoma cell lines by l1-l2 regularization and data reduction. Journal of biomedicine & biotechnology. 2010; 2010:878709.
51. Fardin P, Barla A, Mosci S, Rosasco L, Verri A and Varesio L. The l1-l2 regularization framework unmasks the hypoxia signature hidden in the transcriptome of a set of heterogeneous neuroblastoma cell lines. BMC genomics. 2009; 10:474.
52. Barla A MS, Rosasco L, Verri A. A method for robust variable selection with significance assessment. Proceedings of ESANN. 2008.
53. Jurman G, Merler S, Barla A, Paoli S, Galea A and Furlanello C. Algebraic stability indicators for ranked lists in molecular profiling. Bioinformatics. 2008; 24(2):258-264.
54. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ and von Mering C. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic acids research. 2015; 43(Database issue):D447-452.
55. Wang J, Duncan D, Shi Z and Zhang B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. Nucleic acids research. 2013; 41(Web Server issue):W77-83.
56. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature genetics. 2000; 25(1):25-29.
57. Kanehisa M and Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research. 2000; 28(1):27-30.
58. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K and Yamanaka S. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell. 2007; 131(5):861-872.

59. Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R, Slukvin, II and Thomson JA. Induced pluripotent stem cell lines derived from human somatic cells. Science. 2007; 318(5858):1917-1920.
60. Gratwohl A, Pasquini MC, Aljurf M, Atsuta Y, Baldomero H, Foeken L, Gratwohl M, Bouzas LF, Confer D, Frauendorfer K, Gluckman E, Greinix H, Horowitz M, Iida M, Lipton J, Madrigal A, et al. One million haemopoietic stem-cell transplants: a retrospective observational study. The Lancet Haematology. 2015; 2(3):e91-100.
61. Thorsteinsdottir U, Sauvageau G and Humphries RK. Enhanced in vivo regenerative potential of HOXB4-transduced hematopoietic stem cells with regulation of their pool size. Blood. 1999; 94(8):2605-2612.
62. Pan G and Thomson JA. Nanog and transcriptional networks in embryonic stem cell pluripotency. Cell research. 2007; 17(1):42-49.
63. Nichols J, Zevnik B, Anastassiadis K, Niwa H, Klewe-Nebenius D, Chambers I, Scholer H and Smith A. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell. 1998; 95(3):379-391.
64. Young RA. Control of the embryonic stem cell state. Cell. 2011; 144(6):940-954.
65. Graham V, Khudyakov J, Ellis P and Pevny L. SOX2 functions to maintain neural progenitor identity. Neuron. 2003; 39(5):749-765.
66. Hochedlinger K and Jaenisch R. Induced Pluripotency and Epigenetic Reprogramming. Cold Spring Harbor perspectives in biology. 2015; 7(12).
67. Giorgetti A, Montserrat N, Aasen T, Gonzalez F, Rodriguez-Piza I, Vassena R, Raya A, Boue S, Barrero MJ, Corbella BA, Torrabadella M, Veiga A and Izpisua Belmonte JC. Generation of induced pluripotent stem cells from human cord blood using OCT4 and SOX2. Cell stem cell. 2009; 5(4):353-357.
68. Ito K, Bernardi R and Pandolfi PP. A novel signaling network as a critical rheostat for the biology and maintenance of the normal stem cell and the cancer-initiating cell. Current opinion in genetics & development. 2009; 19(1):51-59.
69. Yilmaz OH, Valdez R, Theisen BK, Guo W, Ferguson DO, Wu H and Morrison SJ. Pten dependence distinguishes haematopoietic stem cells from leukaemia-initiating cells. Nature. 2006; 441(7092):475-482.
70. Guo W, Lasky JL, Chang CJ, Mosessian S, Lewis X, Xiao Y, Yeh JE, Chen JY, Iruela-Arispe ML, Varella-Garcia M and Wu H. Multi-genetic events collaboratively contribute to Pten-null leukaemia stem-cell formation. Nature. 2008; 453(7194):529-533.

SUPPLEMENTARY

Supplementary table 1. Clinical characteristics of the patients and donors

Group	UPN	Age (years)	Diagnosis	Type of transplant	Months from transplant
Cord Blood Unit	1.01	0	-	-	-
	1.02	0	-	-	-
	1.03	0	-	-	-
	1.04	0	-	-	-
	1.05	0	-	-	-
	1.06	0	-	-	-
	1.07	0	-	-	-
	1.08	0	-	-	-
	1.09	0	-	-	-
	1.1	0	-	-	-
	1.11	0	-	-	-
	1.12	0	-	-	-
Adult donors	2.01	22	-	-	-
	2.02	28	-	-	-
	2.03	43	-	-	-
	2.04	36	-	-	-
	2.05	42	-	-	-
	2.06	62	-	-	-
	2.07	47	-	-	-
	2.08	21	-	-	-
	2.09	57	-	-	-
	2.1	55	-	-	-
	2.11	25	-	-	-
	2.12	44	-	-	-
	2.13	38	-	-	-
	2.14	45	-	-	-
	2.15	54	-	-	-
Adult pts post UCBT	3.01	29	AML	IB-UCBT	60
	3.02	50	AML	IB-UCBT	48
	3.03	30	ALL	IB-UCBT	4
	3.04	50	AML	IB-UCBT	90
	3.05	44	AML	IB-UCBT	24
	3.06	22	SAA	IB-UCBT	5
	3.07	20	SAA	IB-UCBT	1
	3.08	18	SAA	IB-UCBT	3
	3.09	54	MDS	IB-UCBT	8
	3.1	55	AML	IB-UCBT	13
	3.11	32	AML	IB-UCBT	12
	3.12	18	ALL	IB-UCBT	5
	3.13	18	AML	IB-UCBT	1

Pts post adult HSCT	4.01	36	AML	PBSCT	7
	4.02	62	ALL	PBSCT	3
	4.03	61	CML	PBSCT	2
	4.04	53	AML	PBSCT	1
	4.05	34	AML	PBSCT	18
	4.06	10	ALL	BMT	1
	4.07	7	FA	PBSCT	2
	4.08	9	ALL	PBSCT	1
	4.09	17	AML	PBSCT	1
	4.1	17	AML	PBSCT	5
	4.11	10	ALL	BMT	5
	4.12	6	Thal	PBSCT	4
	4.13	13	CML	PBSCT	2
	4.14	2	Thal	BMT	3
	4.15	2	Thal	BMT	4
	4.16	3	ALL	PBSCT	1
	4.17	5	ALL	PBSCT	3
	4.18	15	AML	PBSCT	2
	4.19	8	AML	PBSCT	1
	4.2	5	MDS	PBSCT	1
	4.21	2	AML	PBSCT	1
	4.22	7	AML	PBSCT	6
	4.23	15	AML	PBSCT	3
	4.24	11	ALL	PBSCT	1
	4.25	5	AML	PBSCT	1
	4.26	6	Thal	PBSCT	1
	4.27	2	ALL	BMT	6
	4.28	10	ALL	PBSCT	3
	4.29	2	ALL	PBSCT	1

Abbreviations: ALL, acute lymphoid leukemia; THAL, thalassemia; AML, acute myeloid leukemia; MDS, myelodysplastic syndrome; FA, Fanconi anemia; CML, chronic myeloid leukemia; SAA,severe anaplatic anemia.

Supplementary table 2

Gene	Assay code	Antibodies for IF	Gene function
GUSB	Hs99999908_m1		Housekeeping gene
18S	Hs99999901_s1		Housekeeping gene
ABL1	Hs00245445_m1		Housekeeping gene
AXIN1	Hs00394718_m1		Wnt signaling, beta-cateninin-mediated signaling
BCL2L11	Hs00708019_s1		BCL-2 family members, initiator of apoptosis
BMI1	Hs00180411_m1		Repressive activity of many genes, including Hox genes,chromatin remodelling
BMP4	Hs00370078_m1		Resoderm induction
CD44	Hs01075861_m1		Lymphocyte activation, recirculation and homing, and in hematopoiesis
CDKN1A	Hs00355782_m1		Interactor of p53/TP53, regulates proliferation in response to DNA damage
CDKN2A	Hs00923894_m1		Interactor of p53/TP53, regulates proliferation in response to DNA damage
CDKN2C	Hs00176227_m1		Controls cell cycle G1 progression, regulate spermatogenesis
COQ7	Hs01029186_m1		Regulation of metabolism, lifespan determination
CSF2RB	Hs00166144_m1		Cell differentiation
CTNNB1	Hs00355049_m1		Wnt pathway. Stem cell maintenance and proliferation
CXCL12	Hs00171022_m1	SDF-1 (P-159X) : sc-74271	Embryogenesis, immune surveillance, inflammation response, tissue homeostasis
CXCR4	Hs00607978_s1		Leukocyte trafficking, regulation of innate and adaptive immunity

DHH	Hs00368306_m1		Morphogenesis and gonadal development
DLL1	Hs00194509_m1		Differentiation of progenitor cells into the B-cell lineage, emergence of T cell NK cell precursors
DLL3	Hs01085096_m1		Inhibitor of primary neurogenesis
DLL4	Hs00184092_m1		Angiogenesis; inhibitor of endothelial cell proliferation and migration
DPPA2	Hs00414521_g1	DPPA2 (D-17): sc-69392	Maintenance of stem cell pluripotency
DPPA3	Hs01931905_g1		Maintenance of stem cell pluripotency
DPPA4	Hs00216968_m1		Maintenance of stem cell pluripotency
ERCC1	Hs01012158_m1		DNA repair
FGF4	Hs00173564_m1		Embryonic development, cell growth, tissue repair, morphogenesis
FLT3	Hs00174690_m1		Apoptosis, proliferation, and differentiation of hematopoietic cells
FOS	Hs00170630_m1		Cell proliferation, differentiation, and transformation
GDF3	Hs00220998_m1		Cell growth and differentiation in both embryonic and adult tissues
GFI1	Hs00382207_m1		Cell cycle regulator, hematopoietic differentiation
GLI1	Hs00171790_m1		Stem cell proliferation
GLI2	Hs01119974_m1		Embryogenesis
GSK3A	Hs00997938_m1		Cell division, proliferation, motility and survival
HES1	Hs00172878_m1		Regulator of myogenesis
HIF1A	Hs00936371_m1		Regulator of cellular and systemic homeostatic response to hypoxia
HOXA5	Hs00430330_m1		Expression, morphogenesis, and differentiation
HOXA7	Hs00600844_m1		Cell development

HOXA9	Hs00365956_m1		Cell development
HOXB3	Hs00231127_m1	HoxB3 (H-50): sc-28606	Cell development
HOXB4	Hs00256884_m1	HoxB4 (A-15): sc-28607	Stem cell expansion
HOXB5	Hs00357820_m1		Lung and gut development
HOXB7	Hs00270131_m1		Cell proliferation and differentiation
IHH	Hs01081801_m1		Cell growth, patterning and morphogenesis
IL3	Hs00174117_m1		Cell differentiation
IL3RA	Hs00608141_m1		IL3 signaling, activation of STAT pathway
ITGA4	Hs00168433_m1		Cell growth, division, survival, differentiation, migration and apoptosis
ITGB1	Hs00559595_m1		Cell growth, division, survival, differentiation, migration and apoptosis
JAG1	Hs01070032_m1		Notch signaling, cell-fate decisions during hematopoiesis
JAG2	Hs00171432_m1		Embryonic development
KDM5B	Hs00366783_m1		DNA demethylation
KIT	Hs00174029_m1		Regulation of cell survival and proliferation, hematopoiesis, stem cell maintenance
KITLG	Hs00241497_m1		Cell survival and proliferation, hematopoiesis, stem cell maintenance
KLF10	Hs00921811_m1		Regulation of cell growth
KLF2	Hs00360439_g1		T-cell trafficking
KLF4	Hs00358836_m1		Embryonic stem cells maintenance
LIF	Hs00171455_m1		Hematopoietic differentiation
LIG4	Hs00172455_m1		DNA repair
LIN28	Hs00702808_s1	Lin28 (6D1F9):	Maintenance of ES cells

		sc-293120	
MCL1	Hs00766187_m1		Regulation of apoptosis
MCL1	Hs03043899_m1		Regulation of apoptosis
MSH2 DNA	Hs00953523_m1		Mismatch repair system
MYC	Hs00905030_m1		Cell cycle progression, apoptosis and cellular transformation
NANOG	Hs02387400_g1	Nanog (H-2): sc-374103	Early embryogenesis
NES	Hs00707120_s1	Nestin (5C93): sc-71665	Regulation of differentiation, proliferation and apoptosis, lymphoid maturation
NOTCH1	Hs01062014_m1		Regulation of differentiation, proliferation and apoptosis
NOTCH2	Hs01050719_m1		Regulation of differentiation, proliferation and apoptosis
NOTCH3	Hs01128541_m1		Regulation of differentiation, proliferation and apoptosis
NOTCH4	Hs00965889_m1		Regulation of differentiation, proliferation and apoptosis
PCGF2	Hs00810639_m1		Negative regulation of the self-renewal activity of hematopoietic stem cells
PIWIL1	Hs01041737_m1		Stem cell self-renewal, RNA silencing
POU5F1	PHs00999634_Gh	Oct4 (C-10): sc-5279	Early embryogenesis and embryonic stem cell pluripotency
PTEN	Hs02621230_s1	PTEN (B-1): sc-133197	Antagonizes the PI3K-AKT/PKB signaling pathway, neurogenesis
RAD50	Hs00990023_m1		DNA repair
RB1	Hs00153108_m1		Regulator of entry into cell division

RUNX1	Hs01021970_m1		Development of normal hematopoiesis
SHH	Hs00179843_m1		Hedgehog pathway, embryogenesis, stem cell proliferation
SIRT1-	Hs01009006_m1		Ageing
SMAD5	Hs00195437_m1		Inhibition of proliferation of hematopoietic progenitor cells
SMAD7	Hs00998193_m1		Regulator of TGFb signaling, regulator cell cycle arrest in hematopoietic cells
SMARCC1	Hs00268265_m1		Chromatin remodeling, neural development
SMO	Hs01090242_m1		Hedgehog pathway, embryogenesis, stem cell proliferation
SOX1	Hs01057642_s1	Sox-1 (L-20): sc-17317	Embryonic development and in the determination of the cell fate
SOX2	Hs01053049_s1	Sox-2 Ab (S-15): sc-54517	Early embryogenesis and for embryonic stem cell pluripotency
SPI1	Hs02786711_m1		Myeloid and B-lymphoid cell development
STAT3	Hs01047580_m1		Transcription factor, signal transducer , regulators of cell growth and apoptosis
STAT5A	Hs00234181_m1		Transcription factor, signal transducer , regulators of cell growth and apoptosis
TCF3	Hs01012685_m1		Transcription factor regulating B and T lymphocyte development
TGFB1	Hs00998133_m1		Proliferation, differentiation, adhesion, migration
TGFB2	Hs00234244_m1		Proliferation, differentiation, adhesion, migration
TGFBR1	Hs00610318_m1		Regulator of TGFb signaling, regulator cell cycle arrest in hematopoietic cells
TLE1	Hs00270768_m1		Transcriptional corepressor, regulator of Wnt signal

TRIM27	Hs00179059_m1	Enhancer of polycomb protein and repressor of gene transcription, differentiation of germ cells
VPS72	Hs00195618_m1	DNA repair, acethylation, apoptosis, regulation of long term hematopoiesis
XRCC5	Hs00221707_m1	DNA repair
XRCC6	Hs00995282_g1	DNA repair