FINNISH RED CROSS BLOOD SERVICE
AND FACULTY OF BIOCIENCES, DEPARTMENT OF BIOLOGICAL
AND ENVIRONMENTAL SCIENCES, DIVISION OF GENETICS,
UNIVERSITY OF HELSINKI, FINLAND

TRANSCRIPTOME AND GLYCOME PROFILING
OF HUMAN HEMATOPOIETIC
CD133⁺ AND CD34⁺ CELLS

Heidi Anderson

ACADEMIC DISSERTATION

To be publicly discussed, with permission of
the Faculty of Science of the University of Helsinki
in the Nevanlinna Auditorium of
the Finnish Red Cross Blood Service
Kivihaantie 7, Helsinki, on February 15th, at 12 noon.

Helsinki 2008
ACADEMIC DISSERTATIONS FROM THE FINNISH RED CROSS BLOOD SERVICE NUMBER 52

SUPERVISORS

Docent Taina Jaatinen, PhD
Finnish Red Cross Blood Service
Helsinki, Finland

Docent Jukka Partanen PhD
Finnish Red Cross Blood Service
Helsinki, Finland

REVIEWERS

Docent Heli Skottman, PhD
University of Tampere
Regea Institute for Regenerative Medicine
Tampere, Finland

Docent Timo Tuuri, PhD
Infertility Clinic
The Family Federation of Finland
Helsinki, Finland

OPPONENT

Professor, Director Riitta Lahesmaa, MD, PhD
Turku Centre for Biotechnology
University of Turku,
Turku, Finland

ISBN 978-952-5457-17-9 (pdf)
ISSN 1236-0341
http://ethesis.helsinki.fi
Helsinki 2008
Yliopistopaino
To my family
# CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>PUBLICATIONS</td>
<td>6</td>
</tr>
<tr>
<td>ABBREVIATIONS</td>
<td>7</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>8</td>
</tr>
<tr>
<td>REVIEW OF THE LITERATURE</td>
<td>10</td>
</tr>
<tr>
<td>Hematopoietic stem cells</td>
<td>10</td>
</tr>
<tr>
<td>Phenotypic characterization</td>
<td>11</td>
</tr>
<tr>
<td>Functional characterization</td>
<td>14</td>
</tr>
<tr>
<td>Cord blood-derived hematopoietic stem cells</td>
<td>15</td>
</tr>
<tr>
<td>Collection and banking</td>
<td>15</td>
</tr>
<tr>
<td>Cord blood transplantation</td>
<td>16</td>
</tr>
<tr>
<td><em>Ex vivo</em> expansion</td>
<td>21</td>
</tr>
<tr>
<td>Transcriptome profiling</td>
<td>22</td>
</tr>
<tr>
<td>Methods for transcriptome profiling</td>
<td>22</td>
</tr>
<tr>
<td>Gene expression profiling of murine hematopoietic stem cells</td>
<td>26</td>
</tr>
<tr>
<td>Gene expression profiles of human hematopoietic stem cells</td>
<td>28</td>
</tr>
<tr>
<td>Glycosylation</td>
<td>29</td>
</tr>
<tr>
<td>N-glycans and N-glycan synthesis</td>
<td>29</td>
</tr>
<tr>
<td>Glycosylation in hematopoiesis</td>
<td>33</td>
</tr>
<tr>
<td>Effect of glycosylation on homing</td>
<td>33</td>
</tr>
<tr>
<td>AIMS OF THE STUDY</td>
<td>35</td>
</tr>
<tr>
<td>MATERIALS AND METHODS</td>
<td>36</td>
</tr>
<tr>
<td>Ethics</td>
<td>36</td>
</tr>
<tr>
<td>Cells</td>
<td>36</td>
</tr>
<tr>
<td>Methods</td>
<td>37</td>
</tr>
<tr>
<td>Gene expression reanalysis</td>
<td>39</td>
</tr>
<tr>
<td>Annotation</td>
<td>40</td>
</tr>
</tbody>
</table>
RESULTS AND DISCUSSION .......................................................... 41

CD133⁺ and CD34⁺ cells have similar gene expression (studies I and II) .......................................................... 41

Novel molecular markers for HSCs could not be found (study II) .... 42

Differential gene expression between CD133⁺ and CD34⁺ cells suggests that CD133⁺ cells are more primitive (studies I and II) .... 43

CD133⁺ and CD34⁺ cells demonstrate similar total colony forming potential (studies I and II) ........................................................ 44

Reanalysis of the original gene expression data focuses on the similarities between CD133⁺ and CD34⁺ cells ......................... 45

CD133⁺ and CD34⁺ cells have divergent roles in hematopoietic regeneration (studies I and II) .................................................. 49

N-glycosylation-related gene expression is similar in CD133⁺ and CD34⁺ cells and correlates with the N-glycome profile (study III) ... 52

Characteristic mannosylation pattern of CD133⁺ and CD34⁺ cells (study III) ................................................................. 52

Differential expression of N-acetylglucosaminyl-transferase genes and the enrichment of biantennary complex type N-glycans in CD133⁺ cells (study III) ........................................................ 54

Despite the differential expression of genes coding for β1,4-galactosyltransferases there are no differences in the β1,4-galactosylation of CD133+ and CD133- cells (study III) ........................................ 55

CD133⁺ and CD34⁺ cell specific biosynthesis of N-glycan terminal epitopes (study III) .............................................................. 55

CONCLUSIONS ............................................................................ 58

ACKNOWLEDGEMENTS ................................................................ 60

REFERENCES ............................................................................... 62
This thesis is based on the following original publications, which have been reproduced with the permission of the copyright holders.


In addition, some unpublished data are presented.
ABBREVIATIONS

- **BFU** burst-forming erythroid
- **BM** bone marrow
- **CB** cord blood
- **CD** cluster of differentiation
- **cDNA** complementary deoxyribonucleic acid
- **CFU** colony-forming unit
- **DNA** deoxyribonucleic acid
- **E** erythroid
- **GEMM** granulocyte-erythroid-macrophage-megakaryocyte
- **GM** granulocyte-macrophage
- **HSC** hematopoietic stem cell
- **Lin** lineage differentiation
- **N-glycan** asparagine-linked glycan
- **NMR** nuclear magnetic resonance
- **mHSC** murine hematopoietic stem cell
- **mRNA** messenger ribonucleic acid
- **PB** peripheral blood
- **RNA** ribonucleic acid
- **SAGE** serial analysis of gene expression
- **qRT-PCR** quantitative real-time reverse transcriptase-polymerase chain reaction
ABSTRACT

Human blood contains a large variety of differentiated cells with a limited half-life. New blood cells are continuously provided by self-renewing multipotent hematopoietic stem cells (HSC). The capacity of HSCs to regenerate the hematopoietic system is utilized in the treatment of patients with hematological malignancies. HSCs can be obtained from bone marrow (BM) or peripheral blood (PB) after cytokine-induced HSC mobilization. In addition, blood from both placenta and umbilical cord is also rich in HSCs. Cord blood (CB) can be collected after delivery, cryopreserved for the possible future need of an HSC graft for patients with no suitable BM donor available.

HSCs can be enriched using an antibody-based recognition of CD34 glycoprotein on the cell surface. Another glycoprotein, CD133, has also been used for identification of primitive cells. The CD133+ cell population has been shown to be more homogeneous and contain a higher number of potential HSCs than CD34+ cells. CD133 and CD34 molecules are often coexpressed, yet the CD133+ and CD34+ cells may have partly different roles in hematopoiesis.

Each cell has a glycome typical for that cell type. The glycome of HSCs is poorly known. Glycans at the outermost surface of the cell transmit various signals between the matrix environment and the surrounding cells. A CD34+ cell-specific glycoform of CD44, cell adhesion molecule, has been shown to have a possible function in HSC homing.

The aim of present studies has been to characterize the global gene expression of CB CD133+ and CD34+ cells. Furthermore, the identically constituted gene expression profiles of CD133+ and CD34+ cells were compared to unravel shared and unique gene expression in CD133+ and CD34+ cells. A prioritization algorithm was used to rank the genes best separating CD133+ and CD34+ cells from their counterparts, CD133- and CD34- cells. Thus, these genes were suggested to play a part in the regulation of primitive hematopoietic cells. CD133+ and CD34+ cells solely expressed several genes encoding transmembrane proteins that could putatively serve to identify HSCs. Moreover, cell type-specific gene expression was demonstrated in both CD133+ and CD34+ cells. These differentially expressed genes were associated with divergent biological processes suggesting differences in the transcriptional regulation of
CD133+ and CD34+ cells. Thus, they may represent differences in the self-renewal, differentiation or homing properties. Moreover, CD133+ cells expressed a higher number of genes associated with HSC maintenance and pluripotency indicating the more primitive nature of CD133+ cells. Furthermore, the gene expression data bank of CD133+ and CD34+ cells may be used to study the pathogenesis of hematological disorders and the development of malignancies.

Another aim of these studies was to identify the characteristic glycosylation pattern of primitive hematopoietic cells and the genes producing the key glycan entities. High mannose type N-glycans were the most prevalent glycan type in CD133+ cells. CD133+ cells had more biantennary complex type N-glycans than CD133- cells, whereas multiantennary complex type N-glycans were fewer in CD133+ cells than in CD133- cells. Moreover, the CD133+ cell N-glycans showed enrichment of terminal α2,3-sialylation. Combined gene expression and glycan profiling allowed the identification of the key genes regulating the N-glycosylation characteristic for CD133+ cells. The critical genes were similarly expressed in CD34+ cells, suggesting a similar glycosylation pattern in CD133+ and CD34+ cells. Knowledge of HSC glycobiology can be used to design therapeutic cells with improved cell proliferation or homing properties. Improved HSC homing would diminish the number of HSCs needed for the therapy of patients, and allow for greater use of CB grafts.
Hematopoietic stem cells

HSCs are small cell populations in adult BM. They have enormous capacity to self-renew and to generate progenitor cells that can differentiate into mature blood cells\textsuperscript{4}. The life span of most white blood cells is from a few hours to a couple of days. The most common cell type in blood, red blood cells, has an average life expectancy of 120 days. Adult humans have tens of billions of red blood cells. Thus, an average of two million new red blood cells are required every second. HSCs produce new cells to balance the loss of cells each day throughout life.

The mammalian hematopoietic tissue derives from the mesoderm layer of the embryo. The HSCs first germinate in the dorsal aorta\textsuperscript{2}. They are also produced in the major embryonic vessels, umbilical and vitelline arteries\textsuperscript{2}. From the aorta, HSCs migrate to yolk sac, and thereafter to liver and BM. In addition, the placenta supports a large stem cell pool during development and is suggested to serve as source of pre-HSCs/HSCs\textsuperscript{3}. The number of HSCs in the placenta and cord veins drops during the trimester of the gestation. Yet, blood from the umbilical cord and placenta of full-term infant provides a rich source of HSCs.

BM is the primary location for HSCs in the adult. In addition, small amounts of HSCs are in both circulation and in the spleen. Most of the HSCs reside in their BM niche in a quiescent, noncycling state. However, to sustain hematopoiesis and to maintain the HSC pool, cell divisions are required. HSCs have been thought to divide both symmetrically and asymmetrically\textsuperscript{4,5}. In symmetric cell division, two identical HSCs are formed. Asymmetric cell division leads to the formation of two different daughter cells, one that is similar to the mother cell and contains long-term repopulation capacity, and a daughter cell that has short-term repopulation capacity. Long-term HSCs have life-long hematopoietic reconstruction capacity. Short-term HSCs have reduced capacity to sustain hematopoiesis, but they give rise to lineage committed progenitors. According to estimates, only 1/10 000 cells in BM has long-term repopulation capacity\textsuperscript{6}. Once HSC has been committed to lineage differentiation, it no longer has HSC capacity and properties.

The tremendous progeny of HSCs is able to repopulate the marrow. HSC transplantation for marrow repopulation after high doses of chemotherapy
or total body irradiation is an established procedure for the treatment of leukemia. HSC transplantations have also been used in therapy for other hematological disorders, immunodeficiency syndromes, inborn errors of metabolism and autoimmune diseases\textsuperscript{7-9}. It has been suggested that HSCs have the capacity to differentiate into cells of nonhematopoietic tissue as well\textsuperscript{10-12}. Current studies are targeted to find out if HSCs can be utilized in the treatment of such diseases as diabetes, Parkinson’s disease and spinal cord injury.

Phenotypic characterization

HSCs are morphologically similar to leukocytes. They are found among small, rounded cells that have little cytoplasm\textsuperscript{13}. HSCs are mainly noncycling. They stain poorly with vital dyes such as rhodamine123, which locates the mitochondria of living cell, or Hoechst stains 33342/33258, which recognizes DNA of live or fixed cells. Poorly stainable noncycling cells are called Rho/Hoechst 33342\textsuperscript{low}, in contrast to Rho/Hoechst 33342\textsuperscript{high} which indicate the actively cycling cell population. Thus, Rho\textsuperscript{low} cells contain a higher number of HSCs than Rho\textsuperscript{high} cells\textsuperscript{14}. Moreover, HSCs have high aldehyde dehydrogenase activity that can be used to enrich primitive cell populations\textsuperscript{15}.

Leukocytes carry various cluster of differentiation (CD) molecules on their cell surface. These have been used as markers to characterize cell type, maturity and activity of the cell. There is no single marker to identify HSCs, but the presence of the CD34 molecule marks a highly enriched HSC population. As a result, CD34-based isolation of HSCs is most widely used. CD34 antigen is also carried by hematopoietic progenitor cells and cells in the vascular endothelium. Determination of the cell type depends on which other CD molecules are present with CD34. HSCs are found among a population of CD34\textsuperscript{+} cells that do not express CD38 or other lineage differentiation (Lin) markers such as CD2, CD3, CD14, CD16, CD19 CD24, CD56, CD66b and Glycophorin A (Figure 1).
Figure 1. Schematic representation of the hematopoietic lineage differentiation markers during hematopoiesis. Modified after Kyoto Encyclopedia of Genes and Genomes Pathway Database\textsuperscript{16}.
However, other markers such as CD90 (Thy-1) and CD117 (c-Kit) are carried on HSCs. The expression of CD135 (FLT3) is typical for multi-potent progenitors\textsuperscript{17}.

Hematopoiesis has been studied mainly in the mouse model. Murine HSCs (mHSC) have phenotypic characteristics divergent of human HSCs. They are enriched by selecting Lin\(^-\) cells that carry Sca-1 and c-Kit molecules on their cell surface\textsuperscript{18}. Unlike in humans, mHSCs are CD34 negative\textsuperscript{18,19}. CD90 is expressed only at low level in mHSCs\textsuperscript{20}. The signaling lymphocyte activation molecule family receptors, the presence of CD150 molecule and the lack of CD48 molecule, can also be used to highly purify mHSCs\textsuperscript{21}.

A small population of CD34\(^-\)CD38\(^-\) cells carrying the CD133 molecule on their surface seems to be able to give rise to CD34\(^+\) cells in humans\textsuperscript{22}. However, in human HSCs, long-term marrow repopulating capacity is obtained only after the expression of CD34\textsuperscript{23}. On average, 80% of the CD34\(^+\) cells carry CD133 on their surface. CD133\(^+\) cells are almost all CD34 positive as well\textsuperscript{24}. Adult stem cells of non-hematopoietic origin have been found to express CD133\textsuperscript{25-27}. Its expression is also typical in the embryoid bodies derived from embryonic stem cells\textsuperscript{28}. The CD133 molecule is present in circulating cells with endothelial differentiation capacity, suggesting that the differentiation capacity of CD133\(^+\) cells might be greater than in CD34\(^+\) cells\textsuperscript{25}.

HSC enrichment based on antibody-recognition of cell surface molecules is not an ideal method. Antibody binding on HSCs decreases HSC engraftment potential, which is an outcome that should be avoided\textsuperscript{29}. HSC markers have been present on functionally limited cells\textsuperscript{30}. HSCs can have altered cell surface marker expression in \textit{in vitro} conditions\textsuperscript{31}. Furthermore, both functionally competent HSCs and cells with less capacity may have the same phenotype. Many ongoing investigations are aimed toward HSC enrichment by methods that take use of such methods as density gradient difference between cell types or the divergent selectin binding affinity of HSCs\textsuperscript{32-35}. These methods may exceed up to a three-fold enrichment of HSCs when compared to a mononuclear cell population. However, no methods have been more efficient in HSC identification than antibody-based strategies. In the immunomagnetic selection of CD34\(^+\) or CD133\(^+\) cells purities of over 90% can be obtained\textsuperscript{36}. Evaluation of HSC potential requires the study of their functional properties.
Functional characterization

Functional assays are useful in the measurement of the potential of the cell. However, they are complicated to use for cell selection purposes. Better understanding of HSC biology may help to develop novel methods for fast read-out of HSC potential.

The best model used to measure long-term marrow repopulating capacity is provided by severe combined immunodeficiency mice repopulating assay. Of the severe combined immunodeficiency mice transplanted human CD34+CD38- cells, one out of 617 cells is able to engraft. The relatively low capacity of the CD34+CD38- cell population to repopulate mouse marrow is affected by the xenoenvironment. Engraftment capacity has increased with direct intra-BM transplantation, where one out of 44 CD34+CD38- cells has been able to populate the marrow. Human CD34+ cells negative for CD133 have not been able to provide engraftment in mice model. Moreover, the number of vacant HSC niches in the recipient’s BM regulate the engraftment of transplanted HSCs.

In vitro assays, such as long-term culture initiating cells or cobble stone area forming cells, have also been used to estimate the number of long-term HSCs. In these methods, stromal cells are used to support HSC development and their colony forming capacity is tested after the long-term culture. However, differential kinetics between cells in in vivo and in vitro assays indicate that long-term culture conditions may not necessarily be able to represent the in vivo long-term repopulation capacity of HSCs.

Colony forming unit assay (CFU) exploits a cell’s ability to produce progeny in vitro. The cells possessing multipotent progeny, the highest differentiation potential, form granulocyte erythrocyte macrophage megakaryocyte colonies. CD133+CD34+ cells’ progeny are mainly granulocyte macrophage and granulocyte erythrocyte macrophage megakaryocyte colonies. Thus, CD133+ cells are suggested to be more naive than CD133-CD34+ cells, whose progeny also produce some unipotent burst-forming erythrocyte and erythrocyte colonies.
Cord blood-derived hematopoietic stem cells

CB CD34+ cells have greater proliferation potential and they produce significantly more progeny than BM CD34+ cells\textsuperscript{10,42,43}. Thus, the greater number of more immature HSCs may be found in CB than in BM origin\textsuperscript{44}. Moreover, CB HSCs have been shown to be able to differentiate into endothelial and cardiomyocyte-like cells\textsuperscript{45}. Unlike BM CD34+ cells, the CB CD34+ cell population contains very few B cell lineage committed CD34+CD19+ progenitors (Figure 1). In addition, late-erythroid CD34+CD71\textsuperscript{high}CD64-, late-myeloid CD34+CD33+CD15+ and granulomonocytic CD34+CD71+CD64+ progenitors are rarer in CB than BM\textsuperscript{46}. On the other hand, early-myeloid CD34+CD33+CD15- and T cell lineage committed CD34+CD7+CD3- cells are more abundant in CB than in BM.

In cell culture, CB-derived CD34+ cells significantly increase their proliferation when exposed to external cytokines and growth factors and are easier to culture \textit{in vitro} than BM CD34+ cells\textsuperscript{47,48}. The environment of CB HSCs is different from that of BM HSCs. BM HSCs are attached to the osteoblastic cells via the engagement of very late antigen 4 and vascular cell adhesion molecule 1\textsuperscript{49,50}. The osteoblastic cells secrete growth factors and cytokines such as interleukin-6, stem cell factor and chemokine (C-X-C motif) ligand 12, and provide ligands such as fibronectin for the adhesion of HSC\textsuperscript{51}. CB HSCs are, in contrast, devoid of osteoblastic cells. While the placenta is a supportive niche for HSCs in the fetal environment during midgestation\textsuperscript{3}, the placenta and CB of a full-term infant is likely to contain low amounts of growth factors and cytokines. Moreover, CB HSCs may be naturally prepared for oxidative stress and survive better in the 20% oxygen of most cell culture. The oxygen concentration in the CB environment is higher than in BM endosteum, the niche for BM HSCs\textsuperscript{51}. The greater success in culturing CB HSCs may also be augmented by the fact that CB HSCs have markedly longer telomeres than in BM or PB HSCs\textsuperscript{52}. This suggests that CB HSCs are able to go through more cell duplications than BM HSCs.

Collection and banking

After the first successful CB transplantation in 1988\textsuperscript{53}, CB has been stored and cryopreserved for potential future use. CB banks have been created to increase the availability of CB units. Today, there are tens of public CB banks worldwide storing more than 200,000 CB units\textsuperscript{54,55}. In addition, an unknown number of CB units have been stored privately.
The CB units stored in public banks are available to practically everyone. The International NETCORD Foundation provides information regarding the CB banks jointly in Europe, the United States, Japan and Australia. An online virtual office for real-time searching allows an even faster survey of available CB grafts and makes their use more efficient. This especially enhances the search of grafts for patients with a minority genetic background that have only a few suitable stem cell donors available. The accredited CB banks of the Foundation for the Accreditation of Cellular Therapy, such as the Finnish Red Cross Blood Service Cord Blood Bank, ensure that minimum requirements for CB quality, handling and storage are met. As a result, accreditation criteria maintain the quality of the banked CB.

CB is a byproduct of delivery, it is easy to obtain and ex utero collection carries no risk to the donor. CB collection is similar to PB collection, i.e. the cord vein is punctured with a needle and the collection bag is filled by gravity. Suitable CB donors have no known genetic or transmitted disease, have been informed of the situation and have provided written consent for the utilization of the donated CB. Further, CB is tested for human immunodeficiency virus, hepatitis C virus, hepatitis B virus, human T cell leukemia virus and syphilis. In addition, donor mothers have been analyzed for the presence of parvo and cytomegalovirus.

Yet it is not known whether CB cells, cryopreserved for long periods of time, are successful in transplantation. The results on mouse marrow populating assay and CFU suggest a rather long half-life for stored CB units. Thawed CB HSCs have been able to repopulate the marrow in a mouse model after 15 years of storage in liquid nitrogen. The colony forming potential of HSCs has been shown to be 97% of that of fresh HSCs after 12 years of storage in liquid nitrogen.

Cord blood transplantation

Because fast treatment of the patient may be crucial, CB is a noteworthy alternative to transplantation when no suitable related BM donor is readily available. CB as readily available and cryopreserved in CB banks is faster and easier to obtain than unrelated BM donor graft. To obtain BM graft, the donor candidates in the BM donor registry have to first be located. Their HLA-typing results are confirmed, they are informed of the collection process and their consent is requested. Moreover, donors need to pass a complete physical examination and and need to take
additional blood tests to become a suitable BM donor. After all this, the date for transplantation can be scheduled. An average of 13.5 or 49 days are required to obtain CB or BM graft, respectively.\textsuperscript{60}

CB transplantation has already been performed on approximately 4000 patients. CB has been used to treat a variety of diseases\textsuperscript{7-9,55,61}. A full list of diseases treated by CB transplantation is shown in Table 1. Novel applications aim to improve the engraftment of CB grafts. Methods such as HSC expansion, supplement of CD34\textsuperscript{+} cells from PB, mesenchymal stem cells and platelet microparticles are under investigation\textsuperscript{62-67}. Gene transfer into CB HSCs has been used to improve HSC capacity and their use in non-hematopoietic injured tissues. Initial transduced CD34\textsuperscript{+} cells have been successfully used in the treatment of children with severe combined immune deficiency caused by adenosine deaminase-deficiency\textsuperscript{68}. These patients have been given autologous CB CD34\textsuperscript{+} cells transduced with an adenosine deaminase gene.

Human leukocyte antigen-matched adult donors are not available for all the patients. A patient receiving an unrelated CB graft has an equal survival prognosis than patients getting a BM graft from an unrelated donor\textsuperscript{69}. However, patients receiving a human leukocyte antigen-mismatched CB graft from an unrelated donor show less severe graft versus host disease than patients treated with an human leukocyte antigen-mismatched BM or PB graft\textsuperscript{70-75}. It has been suggested that greater tolerance of human leukocyte antigen-disparity in CB transplantation is due to the immature lymphocytes of newborn\textsuperscript{76}. Also, CB dendritic cells are more immature and have diminished functionality when compared to PB\textsuperscript{73}. Dendritic cells present antigens to T and B cells and are responsible for initiating and shaping the adaptive immune system.

The main inconvenience of CB transplantation is the limited number of cells available in a single CB unit. In CB transplantation, 2x10\textsuperscript{7} cells per kilogram is typically considered adequate for transplantation\textsuperscript{77}. The number of cells per kilogram is 10 times more in BM transplantation. Lower cell dose causes less graft versus host disease, whereas higher cell dose has been associated with faster myeloid and platelet engraftment\textsuperscript{71}.

CB grafts have been mostly used in the treatment of pediatric patients. However, reduced intensity transplantation or double CB grafts have been incorporated to overcome the shortage of cells for the therapy of adult patients\textsuperscript{78-80}. Smaller cell doses are suitable for reduced intensity
transplantation which may be applied when patients are not strong enough to go through standard pre-transplant treatment to destroy disease cells\textsuperscript{81}. In the reduced intensity transplantation, the donor immune cells are used to fight the disease. Moreover, the double CB grafts, when compared to the single CB grafts, have shown better engraftment and improved graft versus leukemia response in patients\textsuperscript{82,83}.

Despite the lower risk of graft versus host disease with a CB graft, a similar graft versus leukemia response has been seen after CB and BM transplantations. Immune reconstruction, recovery of early T cells and B cells after CB transplantation, is similar to that of BM\textsuperscript{85,86}. However, some clinics have observed an increase in treatment-related mortality with CB grafts. This has been suggested to be due to slower engraftment of the graft, but could also be due to greater human leukocyte antigen-disparity or lower cell dose of CB grafts used for the treatment. Slower engraftment increases the risk of bacterial infections, which seem to be slightly more common after CB transplantation\textsuperscript{87}. However, infection-related mortality is not increased after CB transplantation\textsuperscript{87}.
**Table 1.** List of diseases and conditions treated with CB transplantation. Adapted from Cord Blood Center⁸⁴.

<table>
<thead>
<tr>
<th>Disease Category</th>
<th>Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Acute leukemias</strong></td>
<td>Acute Biphrenotypic leukemia</td>
</tr>
<tr>
<td></td>
<td>Acute lymphocytic leukemia</td>
</tr>
<tr>
<td></td>
<td>Acute myelogenous leukemia</td>
</tr>
<tr>
<td></td>
<td>Acute Undifferentiated leukemia</td>
</tr>
<tr>
<td><strong>Chronic leukemias</strong></td>
<td>Chronic lymphocytic leukemia</td>
</tr>
<tr>
<td></td>
<td>Chronic myelogenous leukemia</td>
</tr>
<tr>
<td></td>
<td>Juvenile myelomonocytic leukemia</td>
</tr>
<tr>
<td><strong>Myelodysplastic syndromes</strong></td>
<td>Amyloidosis</td>
</tr>
<tr>
<td></td>
<td>Chronic myelomonocytic leukemia</td>
</tr>
<tr>
<td></td>
<td>Refractory anemia</td>
</tr>
<tr>
<td></td>
<td>Refractory anemia with excess blasts</td>
</tr>
<tr>
<td></td>
<td>Refractory anemia with excess blasts in transformation</td>
</tr>
<tr>
<td></td>
<td>Refractory anemia with ringed sideroblasts</td>
</tr>
<tr>
<td><strong>Stem cell disorders</strong></td>
<td>Aplastic anemia (severe)</td>
</tr>
<tr>
<td></td>
<td>Congenital cytopenia</td>
</tr>
<tr>
<td></td>
<td>Dyskeratosis congenital</td>
</tr>
<tr>
<td></td>
<td>Fanconi anemia</td>
</tr>
<tr>
<td></td>
<td>Paroxysmal nocturnal hemoglobinuria</td>
</tr>
<tr>
<td><strong>Myeloproliferative disorders</strong></td>
<td>Acute myelofibrosis</td>
</tr>
<tr>
<td></td>
<td>Agnogenic myeloid metaplasia</td>
</tr>
<tr>
<td></td>
<td>Essential thrombocytopenia</td>
</tr>
<tr>
<td></td>
<td>Polycythemia vera</td>
</tr>
<tr>
<td><strong>Lymphoproliferative disorders</strong></td>
<td>Hodgkin’s disease</td>
</tr>
<tr>
<td></td>
<td>Non-Hodgkin’s lymphoma</td>
</tr>
<tr>
<td></td>
<td>Prolymphocytic leukemia</td>
</tr>
<tr>
<td><strong>Plasma cell disorders</strong></td>
<td>Multiple myeloma</td>
</tr>
<tr>
<td></td>
<td>Plasma cell leukemia</td>
</tr>
<tr>
<td></td>
<td>Waldenström’s macroglobulinemia</td>
</tr>
<tr>
<td><strong>Other malignancies</strong></td>
<td>Brain tumors</td>
</tr>
<tr>
<td></td>
<td>Breast cancer</td>
</tr>
<tr>
<td></td>
<td>Ewing sarcoma</td>
</tr>
<tr>
<td></td>
<td>Neuroblastoma</td>
</tr>
<tr>
<td></td>
<td>Ovarian cancer</td>
</tr>
<tr>
<td></td>
<td>Small-cell lung cancer</td>
</tr>
<tr>
<td></td>
<td>Renal cell carcinoma</td>
</tr>
<tr>
<td></td>
<td>Testicular cancer</td>
</tr>
<tr>
<td><strong>Autoimmune diseases</strong></td>
<td>Even syndrome</td>
</tr>
<tr>
<td></td>
<td>Multiple sclerosis</td>
</tr>
<tr>
<td></td>
<td>Rheumatoid arthritis</td>
</tr>
<tr>
<td></td>
<td>Systemic lupus erythematosus</td>
</tr>
<tr>
<td><strong>Phagocyte disorders</strong></td>
<td>Chediak-Henney syndrome</td>
</tr>
<tr>
<td></td>
<td>Chronic granulomatous disease</td>
</tr>
<tr>
<td></td>
<td>Neutrophil actin deficiency</td>
</tr>
<tr>
<td></td>
<td>Reticular dysgenesis</td>
</tr>
</tbody>
</table>
### Lysosomal storage diseases
- Adrenoleukodystrophy
- Gaucher's disease
- Hunter's syndrome
- Hurler's syndrome
- Krabbe disease
- Maroteaux-Lamy syndrome
- Metachromatic leukodystrophy
- Morquio syndrome
- Mucolipidosis II
- Mucopolysaccharidoses
- Niemann-Pick disease
- Sanfilippo syndrome
- Scheie syndrome
- Sly syndrome, Beta-glucuronidase deficiency
- Wolman disease

### Histiocytic disorders
- Familial erythropagocytic lymphohistiocytosis
- Hemophagocytosis
- Histiocytosis-X
- Langerhans' cell histiocytosis

### Inherited erythrocyte abnormalities
- Beta thalassemia major
- Blackfan-Diamond anemia
- Pure red cell aplasia
- Sickle cell disease

### Congenital (inherited) immune system disorders
- Absence of T & B cells SCID
- Absence of T cells, B cell normal SCID
- Ataxia-telangiectasia
- Bare lymphocyte syndrome
- Common variable immunodeficiency
- DiGeorge syndrome
- Kostmann syndrome
- Leukocyte adhesion deficiency
- Omenn's syndrome
- Severe combined immunodeficiency (SCID)
- SCID with adenosine deaminase deficiency
- Wiskott-Aldrich syndrome
- X-Linked lymphoproliferative disorder

### Other inherited disorders
- Cartilage-hair hypoplasia
- Ceroid lipofuscinosis
- Congenital erythropoietic porphyria
- Glanzmann thrombasthenia
- Lesch-Nyhan syndrome
- Osteopetrosis
- Tay Sachs Disease

### Inherited platelet abnormalities
- Congenital amegakaryocytic thrombocytopenia

### Potential therapeutic applications
- Alzheimer's disease
- Diabetes
- Heart disease
- Liver disease
- Muscular dystrophy
- Parkinson's disease
- Spinal cord injury
- Stroke
Ex vivo expansion

The ex vivo expansion systems have succeeded in augmenting HSC proliferation and increasing the number of progenitor cells with the ability to generate T lymphocytes and Natural Killer cells. The challenge remaining is the expansion of long-term HSCs. CD34\(^+\)CD38\(^-\) cells have been expanded in culture, but the marrow repopulating capacity has not been increased. Moreover, expanded and cycling cells have demonstrated increased susceptibility to apoptosis when compared to quiescent HSCs. This could cause the early death of transplanted HSCs and impaired initial engraftment.

In the ex vivo expansion, the CB-derived cells are cultured in a media supplemented with various growth factors and cytokines such as interleukin-1\(\beta\), -3, -6, -11, stem cell factor, fms-tyrosine kinase 3 ligand, trombopoietin, granulocyte-macrophage colony stimulating factor, erythropoietin and chemokine ligand 3. Alternatively, HSC growth has been supported by co-culture with stromal cells or mesenchymal stem cells. A variation of this method uses conditioned media containing secreted factors from stroma or mesenchymal stem cells.

The engraftment potential of HSCs is limited mostly to cells in the G\(0\) state. HSCs induced to proliferate do not reenter the G\(0\) state. These are the challenges that need to be overcome in order to expand HSCs. Fortunately, the quiescent state of the cell is not as critical to the engraftment potential of CB HSCs as it is to PB HSCs.

A better understanding of molecular mechanisms controlling self-renewal would enhance the ability to expand CB cells. The protein level of transcription factor Homeobox B4 (HOXB4) has been shown to positively correlate with better in vitro proliferation. HOXB4 transduced HSCs have been shown to maintain mHSC capacity. Unfortunately, the same outcome has not been achieved in humans. The potential risk with transduction-based method and cell culture is that they change HSC gene expression and the expression of cell surface molecules. This may have an effect on HSC properties such as homing and therefore on the engraftment of the transplanted cells.
Transcriptome profiling

In recent years, transcriptome profiling has been widely used to understand the genetic regulation of a particular cell type. Transcriptome encompass the global gene expression in a cell at a specific time. It can proffer valuable information on the significant biological processes behind the maintenance of the functionality of the cell. Transcriptome profiling is used to reveal the key genes involved in HSC self-renewal versus differentiation decisions.

The technology for the study of transcriptome is not dependent on any prior knowledge of the genes expressed in the cells of the study. However, challenges remain in regards to the administration and interpretation of the enormous data provided by transcriptome profiling. Transcriptomics provides fundaments for more definitively designed studies and guidance to select the genes for functional studies. It should be kept in mind that the small change in the expression level of a gene may relate to a great difference in the total amount of the corresponding protein present in the cell. The presence of a molecule in a cell depends on the translation features of the mRNA sequence and the amount of antisense sequences degrading the corresponding sense sequences\textsuperscript{95,96}. Results from the transcriptome and proteome analysis of a similar cell type can not usually be fully matched partly due to the heterogeneity or instability of the proteins\textsuperscript{97}. Technical limitations may also explain some of the divergences found in transcriptome and proteome level\textsuperscript{98}.

Methods for transcriptome profiling

Gene expression profiling of 45 \textit{Arabidopsis thaliana} genes was reported in 1995 as the first microarray-based experiment\textsuperscript{99}. Today there are many different kinds of microarray chips used, some of which are the laboratories own cDNA arrays and others which are commercially produced microarrays. The basic idea behind these is the same (Figure 2). Preanalysis and normalization of the data prior to the analysis of the results is one noticeable part of the data interpretation.

Microarray analysis is still the most widely employed technology for transcriptome profiling, but serial analysis of gene expression (SAGE) (Figure 3) and subtractive hybridization (Figure 4) have also been commonly used. Microarray is designed for the study of differential gene
Figure 2. Schematic representation of microarray technology. Total RNA from the cell is isolated and converted into cDNA. cDNA is labeled and hybridized to an array containing oligonucleotide probes. After hybridization, the level of fluorescence of the attached sequences on the chip is measured with a laser scanner, and microarray analysis methods are used to calculate the abundance of each represented RNA. Adapted from Affymetrix GeneChip Expression Analysis Technical Manual.

expression. The subtractive hybridization reveals solely the differentially expressed genes. On the other hand, SAGE technology is efficient in detection of the copy number of the transcripts expressed.
Figure 3. The overview of SAGE technology. RNA from the cells is converted into cDNA molecules in the presence of oligo(dT) magnetic beads. Immobilized cDNA is cleaved with an anchoring enzyme (usually \textit{NlaIII}) and tagged (usually with \textit{BsmFI}) to form SAGE tags. Tags are combined to di-tags, amplified and linked to concatamers of several SAGE tags. Concatamers are cloned in bacteria and then sequenced to reveal the identity and abundance of unique sequences. Modified after (Broxmeyer HE 2004) with data from \textit{Serial Analysis of Gene Expression}\textsuperscript{101,102}.

Several other methodologies such as differential display analysis, representational difference analysis, massive parallel signature sequencing and analysis of expressed sequence tags can also be utilized. The advantages and disadvantages of the three most commonly used...
Figure 4. The general concept of subtractive hybridization. RNA or cDNA from the cells of interest is hybridized to single-stranded cDNA-containing paramagnetic beads from control cells. Hybrids and excess driver cDNA are removed using a magnet. The remaining fragments present the RNA expressed specifically in the cells of interest. Modified from (Lönneborg et al. 1995)\textsuperscript{103}.

technologies; microarray analysis, SAGE and subtractive hybridization are summarized in Table 2. The high costs of using several high throughput methods often limit parallel use of these technologies by a single laboratory. Instead, quantitative real-time reverse transcriptase-
polymerase chain reaction (qRT-PCR) is often used to verify global gene expression profiling results. Usually this means that expressions of a selected set of genes are reanalyzed.

**Table 2.** Comparison of three different methods for gene expression analysis.

<table>
<thead>
<tr>
<th></th>
<th>Microarray</th>
<th>SAGE</th>
<th>Subtractive hybridization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spesificity</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>Moderate</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Detection</td>
<td>Sequences presented on the array</td>
<td>All sequences</td>
<td>All sequences</td>
</tr>
<tr>
<td>Laborated</td>
<td>Moderate</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Need for expertise</td>
<td>Yes</td>
<td>Not much</td>
<td>Not much</td>
</tr>
<tr>
<td>Specific instrumentation</td>
<td>Required</td>
<td>Not required</td>
<td>Not required</td>
</tr>
<tr>
<td>Throughput</td>
<td>High</td>
<td>Moderate</td>
<td>Moderate</td>
</tr>
<tr>
<td>Combarability with other data sets</td>
<td>Possible</td>
<td>Easy</td>
<td>Usually not</td>
</tr>
</tbody>
</table>

It has been shown that results from qRT-PCR, microarray and SAGE are usually correlated, especially when the gene is expressed on the high or medium level\(^\text{104}\). As a sequence-based technology, the sensitivity of qRT-PCR and SAGE analysis is higher than the sensitivity of hybridization-based microarray analysis. Both qRT-PCR and SAGE are able to detect gene expression at a very low level. The level of differential gene expression is often detected lower in microarray analysis than in qRT-PCR\(^\text{105}\).

**Gene expression profiling of murine hematopoietic stem cells**

Previously, particular with the earliest studies, the focus was on discovering transcriptional regulation of mHSCs. Despite the differences in the genes related to hematopoiesis in mice and humans, murine gene expression profiling has proven to be valuable in understanding HSC biology. Genetic programming in stem cells has been underlined in the comparison of the transcriptional expression in HSCs, progenitors and mature hematopoietic cells.
The first large scale analysis of HSCs identified over 2000 genes differentially expressed in fetal mHSCs (AA4.1\(^{+}\)Sca\(^{+}\)c-Kit\(^{+}\)Lin\(^{-}/\text{low}\)) and mature leukocytes (AA4.1\(^{-}\))\(^{106}\). The primitive cells showed enriched expression of genes associated with cell signalling, RNA synthesis, metabolism, protein synthesis, cell division, cell structure, and cell defence. The majority (53%) of the expressed transcripts were not previously known. Gene expression associated with cell signalling has been particularly prevalent in long-term HSC populations, whereas an increase in cell cycle initiation, DNA repair and protein synthesis has been shown to be characteristic of short-term HSCs. Genes typical to lineage committed progenitors are expressed in low levels in short-term mHSCs\(^{107}\).

In the study of gene expression in HSCs and various lineage committed progenitors, 43% of the differentially expressed genes were shown to be specific to the HSC population\(^{108}\). Gene expression profiles of fetal and adult mHSCs have been shown to have 70% overlap\(^{109}\). Fetal HSCs express a few embryonic stem cell-related genes such as Dnmt3b and microH2A1.2 that have not been found to be prevalent in adult BM HSCs. It has also been shown that fetal HSCs are in an active state of cell cycle more often than adult HSCs\(^{110,111}\).

HSCs and neural stem cells have shown common gene expression not present in mature hematopoietic and neuronal cells\(^{108}\). However, highly HSC enriched cells have shown more similar gene expression patterns with mature hematopoietic cells than with stem cells from neural or embryonic origin\(^{112}\). Different stem cell types possess more similarities in the representative biological processes than on the level of individual genes\(^{113}\). Genes associated with chromatin remodeling have been upregulated in various stem cell populations. It has been suggested that this process regulate the self-renewal and differentiation potential of stem cells\(^{114}\). Of the genes common to all stem cell populations, only 4 transcripts uridine phosphorylase, suppressor of Lec15 and two expressed sequence tags have been identified. This suggests that only a few genes have restricted expression in stem cells\(^{112}\).
Gene expression profiles of human hematopoietic stem cells

The differential gene expression between BM, PB and CB-derived CD34+ cells may be related to the possible differing roles of CD34+ cells in different environments\textsuperscript{10,109,115}. The HSC mobilization actuate the gene expression resulting in partly divergent gene expression in BM or granulocyte colony-stimulating factor mobilized CD34+ cells from humans\textsuperscript{116}. Cell cycle promoting genes are more common in BM CD34+ cells than mobilized PB CD34+ cells. The number of genes encoding transcription factors or gene products associated with apoptosis is abundant in mobilized cells when compared to BM CD34+ cells\textsuperscript{116}. The gene expression in CB and BM CD34+ cells is more similar than gene expression in mobilized PB CD34+ cells\textsuperscript{10,115}. BM CD34+ cells have shown enriched expression of genes associated with support of myeloid lineage development and inhibition of apoptosis. Cell adhesion and mobilization-associated genes such as CXCR4, which is important in the homing and engraftment of CD34+ cells, are underexpressed in CB CD34+ cells when compared to BM CD34+ cells.

Typically, a high amount of expressed genes is common to HSCs. This may be due to their relatively open or transcription permissive chromatin structure\textsuperscript{117,118}. CB CD34+ cells have the highest number of overexpressed genes when CB, BM and PB CD34+ cells are compared\textsuperscript{10,115}. The number of overexpressed genes is higher in BM CD34+ cells than in PB CD34+ cells. The greater number of active genes suggests broader transcriptional options for HSCs. Many genes expressed by HSCs have been located in the pericentric heterochromatin, unlike the genes related to activation in mature lymphocytes located in centromeric heterochromatin\textsuperscript{118}. This suggests at least a partly different pathway for the activation of genes in HSC and mature lymphocytes. Its possible role is to provide fast recruiting cells to generate the type of blood cells needed.

Low levels of embryonic stem cell marker gene Oct4 have been detected in CB CD133+ cells\textsuperscript{119}. Other pluripotent markers such as Sox1, Sox2, FGF4 of Rex1 are also expressed in CB CD133+ cells\textsuperscript{119}. These findings suggest that CD133+ cells, especially from CB, may have the potential to differentiate into nonhematopoietic tissue types. Moreover, the slowly dividing CB CD34+CD38- cells have been shown to express several genes differently when compared to fast dividing CD34+CD38- cells\textsuperscript{120}. These slowly dividing, more primitive, hematopoietic cells were shown to express CD133 abundantly.
Glycosylation

Glycosylation is specific to each cell type. It involves common posttranslational modifications that take place on the nascent protein. There are two kinds of glycosylation; N-glycosylation and O-glycosylation. Nearly all proteins traveling through endoplasmic reticulum become N-glycosylated. Glycosylated proteins mainly function in the extracellular compartments and in the secretory pathway. Due to their location on the cell surface, they serve a role as providers of information transfer in cell recognition and in interactions with other cells or in the extracellular matrix.

Revealing glycosylation features characteristic of HSCs broadens the knowledge of HSC biology. Two molecules may have identical amino acid composition, but differ due to differential glycosylation. Diverse glycoforms of Thy-1, a stem cell marker, have been found in the thymus and brain. Moreover, individual glycosylation leads to the formation of blood groups such as ABO(H), Secretor, P and Lewis on erythrocytes. Changes in glycosylation are involved in cell differentiation, embryogenesis and tissue morphogenesis.

N-glycans and N-glycan synthesis

N-glycans support protein folding, stability and oligomerization, and participate in trafficking, recognition of receptor binding sites and protein localization. N-glycans are vital for organisms from yeast to human, and most of the known glycosylation disorders are caused by defective N-glycosylation. The defects have been most dramatic in cells going through rapid alterations, such as leukocytes. For example, the inhibition of N-glycan backbone trimming mannosidases cause a blockage in B cell differentiation. In addition to N-glycan, other glycan types including O-glycans, glycosphingolipids, glycosaminoglycans and glyco-phospholipid anchors are common. Due to the technical challenges of studying the structure of these other glycan types, they are less well known.

In the biosynthetic pathway of N-glycosylation, the preformed glycan is transferred to the amino acid asparagine (N) glycosylation site on the nascent protein in endoplasmic reticulum. N-glycans can be grouped into three classes; high mannose, hybrid and complex type structures, depending on the amount and type of glycan modifications they have gone through (Figure 6).
Figure 5. Schematic representation of N-glycans. N-glycans consist of distinct regions of N-glycan core, backbone and terminal epitopes that are synthesized by different glycosyltransferase and glycosidase families. Monosaccharide symbols: ■ = N-acetyl-D-glucosamine; ○ = D-mannose; □ = D-galactose; △ = L-fucose; ◆ = N-acetylneuraminic acid.

High mannose N-glycans represent the unprocessed glycan structures with a maximum of 9 D-mannoses attached to the N-glycan core. The hybrid type or intermediately processed glycans possess characteristics of both high mannose and complex type N-glycans. Complex type glycans are so-called fully processed N-glycans and occur as bi-, tri-, tetra- or pentaantennary structures.

The α-mannosidase 1 enzymes encoded by MAN1 genes and β-N-acetylglucosaminyltransferase 1 encoded by MGAT1 promotes the conversion of high mannose N-glycans into more processed glycan types (Figure 6). MAN1 gene products remove the terminal α-mannose residues, while β-N-acetylglucosaminyltransferase 1 transfer an N-acetyl-D-glucosamine residue to form the first branch towards processing more developed N-glycans. The synthesis of hybrid and complex type N-glycans is developmentally essential.

MAN2 genes encode α-mannosidase 2 enzymes, which trim the hybrid type N-glycans prior to the action of MGAT2 coded β-N-acetylglucosaminytransferase 2. Further, β-N-acetylglucosaminytransferase 2 forms the complex type N-glycans. Complex type N-glycans have also been found in MAN2-deficient mice despite the absence of family 2 α-mannosidases. This suggests that there are other enzymes, possibly other α-mannosidases, involved in the demannosylation of hybrid type N-glycans. Accumulation of hybrid type N-glycans and the lack of complex N-
glycans in organisms may affect cell signaling. A mutation in the human ortholog of the MGAT2 gene has been associated with a rare human disease known as congenital disorders of glycosylation type IIa. This disease affects disfunction of multiple physiological systems.

Figure 6. N-glycan synthesis. The attachment of N-glycan precursor to the glycosylation site asparagine (N) of a nascent protein initiates the synthesis. Further, the three glucose and one mannose residues are enzymatically removed from the N-glycan. The processing of N-glycan into the hybrid type starts by the removal of the mannose monosaccharides by family 1 mannosidases and adding one N-acetyl-D-glucosamine by β-N-acetylglucosaminyltransferase 1. The rest of the removable mannoses are docked by mannosidase 2 family members. Different glycosyltransferases, modifying the glycan backbone and terminal epitopes, form the vast variety of N-glycans by differentially adding N-acetyl-D-glucosamine, D-galactose, L-fucose and N-acetyleneuraminic acid residues. Monosaccharide symbols:  = glucose;  = N-acetyl-D-glucosamine;  = D-mannose;  = D-galactose;  = L-fucose;  = N-acetyleneuraminic acid. Glycosidic linkages are indicated by lines connecting the monosaccharides.
ated by MGAT4a, MGAT4b and MGAT5 encoded β-N-acetylglucosaminyltransferases. Insufficient expression of MGAT4a has been shown to cause failure in pancreatic beta cells leading to type 2 diabetes\textsuperscript{131}. In some cases, β-N-acetylglucosaminyltransferase 5 forms another branch in the α1,3-linked core mannose. MGAT5 does not seem to be essential for survival\textsuperscript{132}. On the other hand, enhanced MGAT5 expression has been demonstrated to support tumor growth and metastasis\textsuperscript{133}.

The variability of N-glycans is increased with terminal epitopes modified by β1,4-, β1,3-galactosyltransferases, α2,3-, α2,6-sialyltransferases and α1,2-, α3/4-, α1,6-fucosyltransferases (Figure 5). Mutations in the terminal epitopes are not usually crucial for survival, but they often affect cells in the hematopoietic system. Mice with mutations in the B4GALT1 gene have reduced neutrophil trafficking in the inflamed tissues\textsuperscript{134}. In humans, the mutation in B4GALT1 cause congenital disorder of glycosylation type IIId disease with severe neuronal disease and blood clotting defects\textsuperscript{135,136}. The role of the 4 other β1,4-galactosyltransferases in N-glycans synthesis is poorly known. However, it has been suggested that these β1,4-galactosyltransferases have overlapping functions.

Immunodeficiency and attenuated B cell formation has been observed in ST6Gal1-knockout mice\textsuperscript{137}. ST3Gal4 is involved in the biosynthesis of sialyl-Lewis\textsuperscript{x} and sialyl-Lewis\textsuperscript{a} antigens, both known as selectin ligands\textsuperscript{138}. ST3Gal4-deficiency causes inflammation response deficit, formation of deficient platelets and defects in blood coagulation\textsuperscript{139}. The role of ST3Gal6, the fourth α2,3-sialyltransferase in the N-glycan synthesis is still obscure\textsuperscript{140}.

Fucosyltransferases do not seem to be essential in mice\textsuperscript{141-144}. However, FUT genes have an important role in hematopoiesis and the immune response. FUT4-null mice have partial neutrophil adhesion deficit and selectin ligand defect similar to the FUT7-knockout\textsuperscript{144,145}. The selectin ligand activity is completely lost in FUT4 and FUT7 knockout mice\textsuperscript{144}, suggesting that the gene products are pivotal glycosyltransferases responsible for the synthesis of active selectin ligands. In humans, the mutation in the FUT7 gene results in modest reductions of leukocyte selectin ligand activity\textsuperscript{146}.

Considerable glycosylation differences are prevalent between organisms. All other mammals except the humans, apes and Old World monkeys, have functional α1,3-galactosyltransferase\textsuperscript{147}. In humans, the Gal-α1,3-Gal glycoconjugates synthesized by α1,3-Galactosyltransferase
are antigenic and induce serum anti-Gal antibodies\textsuperscript{148}. Moreover, the defect in N-glycolylneuraminic acid synthesis is specific to humans\textsuperscript{149}. The glycosylation differences between organisms need to be taken into account when defining the components of a cell culture media for stem cells to be transplanted into humans\textsuperscript{150,151}.

**Glycosylation in hematopoiesis**

One of the most dramatic alterations of glycosylation occurs at the time of postnatal development, where fetal erythrocyte type polylactosamine changes into branched I type polylactosamine\textsuperscript{152}. Thus, glycans can be used as biomarkers for cellular maturity. Moreover, typical glycosylation features of hematopoietic lineages separate the different cell types. Lymphoid cells have a large amount of N-glycan $\alpha_{2,6}$-sialylation, whereas sialyl-Lewis$^x$ antigens are enriched in myeloid cells\textsuperscript{153,154}. In these cells, a localized increase in $\alpha_{2,6}$-sialylation of CD11b/CD18 molecules has been seen during myeloid cell maturation and liberation from BM\textsuperscript{155}.

Glycosylation of HSCs is still poorly known. $\alpha_{2,6}$-sialylation is also dominant in CD34$^+$ cell N-glycans\textsuperscript{156}. $\alpha_{2,6}$-sialylation has been increased in granulocyte growth stimulating factor mobilized CD34$^+$ cells when compared to naturally circulating CD34$^+$ cells\textsuperscript{157}. It has been suggested that sialylation events modify the adhesion between HSCs and osteoblasts\textsuperscript{158}.

**Effect of glycosylation on homing**

Glycosylation plays an important role in the adhesion and homing of HSCs and leukocytes occurring in a selectin-dependent manner\textsuperscript{159}. Leukocyte rolling on endothelial cells is supported by the interaction between the lymph node homing receptor L-selectin and its ligands. Leukocytes express the P-selectin glycoprotein ligand 1 which is able to bind to L-, P- and E-selectins\textsuperscript{160,161}. The selectin binding of P-selectin glycoprotein ligand 1 is dependent on O-glycosylation\textsuperscript{162}.

CB CD34$^+$ cells have impaired $\alpha_{1,3}$-fucosylation not seen in BM HSCs\textsuperscript{163}. CB CD34$^+$ cells express a form of P-selectin glycoprotein ligand 1 that does not bind to P-selectin\textsuperscript{163}. Enforced $\alpha_{1,3}$-fucosylation of CB CD34$^+$ cells improved the binding to P-selectin possibly in a P-selectin glycopro-
tein ligand 1-dependent manner. α1,3-fucosylation is important for the synthesis of the selectin ligand glycan structure sialyl-Lewisx.

A selectin ligand, CD44, has a specific role in CD34+ cells binding to both L and E-selectin. A CD34+ cell specific glycoform of CD44 contains N-glycosylation, which is crucial to its function. CD44 has been suggested to play a major role in CD34+ cell adhesion and movement along endothelium and leukocytes.

In human leukemia CD34+ KG1a cells, the CD34+ cell-specific CD44 has been shown to have a 5-fold higher affinity to selectin binding than P-selectin glycoprotein ligand 1. Both P-selectin glycoprotein ligand 1 and CD44 may participate in the homing of HSCs and the dysfunctional P-selectin glycoprotein ligand 1 in CB-derived cells may conduce to reduced homing capacity of CB HSCs. The enforced fucosylation of the glycans of CB CD34+ cells have shown contrasting results in their ability to improve their homing.

The modification of HSCs by glycan remodelling may provide new tools for HSC therapeutics. Engineering of cells through glycan modifications could be used to enhance HSC homing capacity. This would provide a greater use for CB HSCs, as a lower number of cells might be sufficient for successful engraftment.
AIMS OF THE STUDY

The aims of this study were

1. To characterize the global gene expression and glycome of human hematopoietic CD133+ and CD34+ cell populations from CB (studies I-III)
2. To reveal the differential gene expression in CD133+ and CD34+ cells and to relate the findings to the characteristic biological processes of these cells (study II)
3. To identify the genes responsible for cell type-specific N-glycosylation in CD133+ and CD34+ cells (study III)
4. To find novel molecular markers specific to HSCs (study II)
MATERIALS AND METHODS

Ethics

This study was approved by the ethical board of the Finnish Red Cross Blood Service and Helsinki University Hospital. The CB donors were informed and written consent was obtained. Only CB units unsuitable for therapeutic use were studied. A small total volume or the low number of total cells were the most common reasons for discharging CB units.

Cells

CB units were collected at the Helsinki Maternity Hospital and the Department of Obstetrics and Gynaecology at Helsinki University Hospital. The collections were performed during 2004-2007. In all the studies, fresh CB units were used.

In study I, CD133+ cells were selected from the 4 CB unit and CD133- cells were collected for control purposes. CD133+ data were compared to CD133- data from the same CB unit in the microarray analysis. The differential gene expression between CD133+ and CD133- cells was to be the same in all replicates. The fold-change cutoff value was set to 3. To verify the information obtained from the microarray data, selected genes were subjected to qRT-PCR analysis using pools of three samples in each. Analysis was performed on two biological replicates. In addition, CD133+ and CD133- cells were used to evaluate the purity of the cell fractions. The CFU assay was performed for CB CD133+, CD133- and mononuclear cells in duplicate.

In study II, CD34+ cells were selected from the 3 CB unit and CD34- cells were collected for control purposes. Similarly to study I, CD34+ data were compared to CD34- data from the same CB unit and the results verified using qRT-PCR analysis. Moreover, the assessment of the purity of the cell fractions and the CFU assay were performed for CD34+ and CD34- cells as in study I. Further, similarly constructed CD133+ and CD34+ gene expression profiles were compared to reveal the specific expression patterns and differences of CD133+ and CD34+ cells. In addition, gene expression data of CD133+ and CD34+ cells were directly compared to the gene expression data of PB mononuclear cells. Thus, the PB mononuclear cell sample was provided as a common
benchmark to control the differential gene expression in CD133+ and CD34+ cells.

In addition, the gene expression profile results suggested several putative membrane proteins that could serve as an HSC marker in the isolation of HSCs. For the analysis of the presence of novel HSC marker molecules, selectively monoclonal antibody-labeled CB mononuclear cells were the subject of the flow cytometric analysis.

In study III, CD133+ and CD133− cell N-glycans were isolated for mass spectrometric analysis. During the preparation process, ultrapure bovine serum albumin was used to avoid oligosaccharide contamination. To demonstrate the quantitative differences between N-glycosylation of CD133+ and CD133− cells, CD133+ and CD133− cell N-glycan profiles were compared to each other. In addition, CB CD133+ and CD8+ T cell N-glycan profiles were compared to each other. The CD8+ T cell N-glycan profile served as an independent, immunogenetically selected mature hematopoietic cell control.

The mononuclear cells were analyzed in nuclear magnetic resonance (NMR) analysis to obtain detailed data about most abundant N-glycan structures. In addition, N-glycan fractions of CD133+ and CD133− cells were digested with specific exoglycosidase enzymes to characterize the terminal epitopes on CD133+ and CD133− cell N-glycans. Further, the glycan profile and glycosylation-related gene expression were compared to unravel the transcriptional regulation of the glycosylation features characteristic of CD133+ cells. Also, the glycosylation-related gene expression of CD34+ cells was observed. The glycan determinants on the cell surface of mononuclear and CD34+ cells were demonstrated by a lectin binding assay. For this, Glycophorin A-depleted CB mononuclear cells from 3 CB units were used.

**Methods**

Methods used in the following studies (I-III) are described in detail in the original publications and are listed in Table 3.
# MATERIALS AND METHODS

## Table 3. The methods used in studies I-III

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Method</th>
<th>Reference</th>
<th>Used in study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isolation of mononuclear cells</td>
<td>Ficoll-Hypaque density gradient centrifugation</td>
<td>Porras et al. 2005</td>
<td>I, II, III</td>
</tr>
<tr>
<td>Mononuclear cell count</td>
<td>Automatic cell counter Sysmex K-1000/Bürker counting chamber</td>
<td>Kekarainen et al. 2006, Nystedt et al. 2006</td>
<td>I, II, III</td>
</tr>
<tr>
<td>Selection of CD133+/CD34+ cells</td>
<td>MiniMACS/MidiMACS immunomagnetic separation system and Midi MACS LS columns</td>
<td>Kekarainen et al. 2006</td>
<td>I, II, III</td>
</tr>
<tr>
<td>Depletion of GlyA+ cells</td>
<td>MidiMACS immunomagnetic separation system and Midi LD columns</td>
<td>Kekarainen et al. 2006</td>
<td>III</td>
</tr>
<tr>
<td>Purity assessment of selected cells</td>
<td>Phycoerythrin- and fluorescein isothiocyanate-conjugated monoclonal antibody labeling and flow cytometric analysis</td>
<td>Kekarainen et al. 2006</td>
<td>I, II, III</td>
</tr>
<tr>
<td>Clonogenic progenitor capacity measurement</td>
<td>Colony-forming unit assay</td>
<td>Kekarainen et al. 2006</td>
<td>I, II</td>
</tr>
<tr>
<td>RNA isolation</td>
<td>RNA isolation kit</td>
<td>RNeasy Mini Handbook</td>
<td>I, II, III</td>
</tr>
<tr>
<td>RNA quality</td>
<td>Spectrophotometric analysis and agarose gel electrophoresis</td>
<td>Wright et al. 1997</td>
<td>I, II, III</td>
</tr>
<tr>
<td>cDNA synthesis and hybridization to the array</td>
<td>Affymetrix sample preparing protocol and Human Genome U133 Plus 2.0 oligonucleotide array</td>
<td>GeneChip Expression Analysis Technical Manual</td>
<td>I, II, III</td>
</tr>
<tr>
<td>Microarray data normalization and analysis</td>
<td>GeneChip Operating Software and statistical analysis</td>
<td>Zou et al. 2005</td>
<td>I, II, III</td>
</tr>
<tr>
<td>Determination of gene expression patterns</td>
<td>Hierarchical clustering and Self-Organizing Map algorithm with U-matrix</td>
<td>Eisen et al. 1998, Hautaniemi et al. 2003</td>
<td>I, II</td>
</tr>
<tr>
<td>Identification of the most differentially expressed genes</td>
<td>A gene selection algorithm</td>
<td>Golub et al. 1999, Brügel et al. 1993</td>
<td>I, II</td>
</tr>
<tr>
<td>Gene annotation</td>
<td>Affymetrix GO Ontology Mining Tool</td>
<td>Zou et al. 2005</td>
<td>I, II, III</td>
</tr>
<tr>
<td>Verification of the microarray results</td>
<td>Quantitative real time-polymerase chain reaction</td>
<td>Kaartinen et al. 2007</td>
<td>I, II</td>
</tr>
<tr>
<td>Detection of novel transmembrane protein expression</td>
<td>Specific fluorescent antibody labeling and flow cytometric analysis</td>
<td>Kekarainen et al. 2006</td>
<td>II</td>
</tr>
<tr>
<td>Isolation of N-glycans</td>
<td>N-glycosidase digestion</td>
<td>Nyman et al. 1998</td>
<td>III</td>
</tr>
<tr>
<td>Analysis of N-glycans</td>
<td>Matrix-assisted laser desorption-ionization time-of-flight mass spectrometry and nuclear magnetic resonance analysis</td>
<td>Aoki et al. 2007, Welkkolainen et al. 2007</td>
<td></td>
</tr>
</tbody>
</table>
Gene expression reanalysis

The improved knowledge on gene and transcript sequences, as well as genome assembly, reveal a great number of previously defined probe sets in Affymetrix microarray GeneChips that are not applicable in the detection of the gene. The unreliable representative gene identifier has been assigned for approximately 18% of the probe sets in the Affymetrix Human Genome U133 Plus 2.0 chip. Moreover, about 12% of the probe sets contain probes that may hybridize to transcripts of more than one gene or to a non-coding sequence, while some probes no longer match any sequence in the current databases. Thus, reanalysis of all the gene expression data was performed to access more accurate annotation of the differentially expressed genes.

Raw data was normalized using the GC Robust Multi-array Average background adjustment for intensities, quantile normalization, and median-polish summarization. Biostatistic tools were used in R software. The probes of each Affymetrix probe set, directed toward identification of a transcript, were redesigned to sets of probes using GeneChip library files (http://brainarray.mbni.med.umich.edu/CustomCDF). The probe sets of these library files were chosen to match a more recently updated Unigene cluster representing certain genes (http://arrayanalysis.mbni.med.umich.edu/ps/). To obtain CD133+ and CD34+ cell gene expression profiles, CD133+/CD34+ cells were compared to CD133-/CD34- cells. From the CD133+ and CD34+ profiles, 500 genes with the most differential expression pattern were selected according to their median fold change. That is, the ratio of the median gene expression level in CD133+/CD133- cells or CD34+/CD34- cells was calculated and transformed to an absolute value of log2. Differentially expressed genes could be overexpressed or underexpressed in CD133+/CD34+ cells when compared to CD133-/CD34- cells. The statistical significance of the differential expression was performed using a T-test. The differentially expressed genes with a q-value>0.05 reflected the probability of a false positive result and were discharged. To directly compare differences in CD133+ and CD34+ cells, a pairwise analysis utilizing GeneSifter software was employed. The significant change in the expression between CD133+ and CD34+ cells was validated using Student’s T-test and a p-value that was ≥0.05.
Annotation

Due to the ongoing changes and revisions in the annotations of Affymetrix probe sets, the differentially expressed genes in studies I and II were reannotated using the GeneAnnot Microarray Gene Annotation software (Weizmann Institute of Science, http://bioinfo2.weizmann.ac.il/geneannot/)[184,185]. The Database for Annotation, Visualization and Integrated Discovery tools were employed to obtain molecular functions, biological processes, and cellular components for the transcripts common in the CD133+ and CD34+ expression profiles in the original and reanalyzed data[186]. The molecular interactions for the protein products were assessed through GeneSifter software, which uses Kyoto Encyclopedia of Genes and Genomes Pathway database[16].
RESULTS AND DISCUSSION

CD133\(^+\) and CD34\(^+\) cells have similar gene expression (studies I and II)

The comparison of CD133\(^+\) and CD133\(^-\) data sets resulted in 690 transcripts that were differentially expressed at least three-fold in all the samples. Of these, 393 transcripts were upregulated and 297 were downregulated in CD133\(^+\) cells. Similarly, CD34\(^+\) data sets were compared to CD34\(^-\) data sets and 620 differentially expressed transcripts with at least a three-fold difference were found. In CD34\(^+\) cells, 169 transcripts were overexpressed and 451 transcripts underexpressed. The integrated sets of differential gene expression between CD133\(^+\) and CD133\(^-\) cells, and CD34\(^+\) and CD34\(^-\) cells, showed that 85\% of the transcripts overexpressed in both CD133\(^+\) and CD34\(^+\) cells and 36\% of the transcripts underexpressed in both CD133\(^+\) and CD34\(^+\) cells were the same. Transcripts found in both CD133\(^+\) and CD34\(^+\) cells showed the highest overexpression when compared to CD133\(^-\)/CD34\(^-\) cells. Both cell types expressed many genes previously identified to be present in HSCs (such as \textit{KIT}, \textit{LAPTM4B}, \textit{MEIS1}, \textit{RBPMS}, \textit{HLF}, \textit{BAALC} and \textit{C17})\(^{109,115,120,187-190}\). Also, many genes supporting the undifferentiated state of HSC (such as \textit{GATA2} and \textit{N-MYC} and \textit{HOXA9})\(^{106,191,192}\). These genes were expressed at a very similar expression level across the CD133\(^+\) and CD34\(^+\) cell replicates indicating an important role for these genes in the regulation of CD133\(^+\) and CD34\(^+\) cells. Both cell types expressed a high number of transcripts. In addition, these transcripts were arranged similarly in SOM analysis suggesting similar regulation for the common genes in CD133\(^+\) and CD34\(^+\) cells. In the CD133\(^+\) profile, the number of overexpressed transcripts was higher than in the CD34\(^+\) profile. This could explain the additional cluster of overexpressed transcripts illustrated in U-matrix of CD133\(^+\) data.

CD133\(^+\) and CD34\(^+\) cells expressed 85 transcripts that were not present in CD133\(^-\) or CD34\(^-\) cells. These transcripts were either downregulated or absent in PB mononuclear cells. The 85 common transcripts represented 80 genes. The encoded molecules are involved in cell proliferation, metabolism and localization. Of these, the expression of 44 genes in HSCs has been confirmed in other studies\(^{109,115,120,188,193}\). The transmembrane protein encoding genes \textit{FLT3 (CD135)}, \textit{ALCAM (CD166)} and \textit{DSG2} were expressed in CD133\(^+\) and CD34\(^+\) cells only. Moreover, expression of \textit{SV2A} was found in CD133\(^+\) cells, but not in CD34\(^+\) cells.
Novel molecular markers for HSCs could not be found (study II)

To determine whether transmembrane protein encoding genes expressed solely in CD133$^+$ and CD34$^+$ cells could provide novel molecular markers for HSCs, the presence of the protein products on the cell surface were observed. Total leukocyte population was labeled with specific antibodies binding to CD135, CD166, DSG2 or SV2A. In addition, anti-CD34 or anti-CD133 antibody labeling was used. Coexpression of CD135 and CD34/CD133 molecules was demonstrated (Figure 7). Moreover, CD166 was expressed on some of the CD133$^+/CD34^+$ cells (Figure 6). CD135 was also present on 11% of the leukocytes and CD166 was carried by 20% of the leukocytes. Thus, neither CD135 nor CD166 alone can be used for selective purification of HSCs.

SV2A or DSG2 molecules could not be found on leukocytes, including CD133$^+/CD34^+$ cells. It is possible that the antibodies against SV2A and DSG2 molecules do not posses adequate specificity. Therefore, no single molecular marker to identify HSCs could be found using microarray analysis. In addition, known embryonic stem cell markers SSEA3, SSEA4, TRA-1-60 and TRA-1-81 were searched for on the leukocytes (data not shown). No hematopoietic cell populations carrying these markers could be demonstrated.

![Figure 7. Presence of CD135 and CD166 molecules on CD34$^+$ and mononuclear cells.](image-url)
Differential gene expression between CD133\(^+\) and CD34\(^+\) cells suggests that CD133\(^+\) cells are more primitive (studies I and II)

Despite the highly similar gene expression in CD133\(^+\) and CD34\(^+\) cells, unique gene expression patterns related to divergent biological processes were distinguished in CD133\(^+\) and CD34\(^+\) cells. The number of overexpressed transcripts in CD133\(^+\) cells, when compared to CD133\(^-\) cells, was higher than in CD34\(^+\) cells when they were compared to CD34\(^-\) cells.

In the study I, a set of 257 transcripts were identified expressed in CD133\(^+\) cells, but not in CD133\(^-\) cells. Of these, 125 genes were also expressed in CD34\(^+\) cells. However, the 172 transcripts overexpressed solely in CD133\(^+\) cells (when compared to CD133\(^-\) cells) showed significant association with regulation of chromatin structure, cell cycle and DNA metabolism. CD133\(^+\) cells expressed \textit{SMARCA1}, \textit{SMARCA2} and \textit{SMARCA4}. \textit{SMARCA} genes are related to chromatin remodeling pathway\(^{194,195}\). Only \textit{SMARCA1} expression was detected in CD34\(^+\) cells. The high number of expressed transcripts is typical for HSCs that have been suggested to have permissive or more open chromatin structure.

Embryonic stem cell-related transcripts including \textit{DNMT3A}, \textit{DNMT3B} and \textit{DPPA4} were shown only in CD133\(^+\) cells. The expression of transcripts related to pluripotency suggests a more primitive nature for CD133\(^+\) cells. The embryonic-type gene expression may also be due to the CB origin of the cells. The expression of \textit{MPDZ}, encoding a protein associated with tight junctions and maintenance of cell polarity, suggests that some of the CD133\(^+\) cells may be going through asymmetric cell division.

The products of the expressed transcripts in CD34\(^+\) cells were related to regulation of development, response to stimulus and DNA metabolism. These biological processes were also typical of CD133\(^+\) cells. On the other hand, CD34\(^+\) cells expressed transcripts encoding markers for lineage committed cells.

Several studies have shown transcriptional as well as functional differences between CD34\(^+\) cells from CB, BM and PB\(^{115,196-198}\). These studies have suggested a more primitive nature for CB CD34\(^+\) cells. Moreover, differential gene expression has been found between CB CD133\(^+\) and PB CD133\(^+\) cells\(^{193}\). This study demonstrated divergent gene expression between CB CD133\(^+\) and CD34\(^+\) cells. The gene expression profiles of
CD133+ and CD34+ cells were identically constructed. The samples were prepared in the same laboratory, hybridized at the same time and analyzed using equal parameters. Therefore, many nonbiological error sources for differential gene expression between the two cell populations were avoided. The CD133+ and CD34+ gene expression profiles of the present study were also compared to CD133+ cell gene expression profile in another study. CD133+ gene expression profiles were more similar to each other than with CD34+ cell gene expression profile of the present study. This confirms that the different transcriptional expression in CD133+ and CD34+ cells is due to differences in these two cell populations.

The differentially expressed transcripts between CD133+ and CD34+ cells were related to divergent biological processes, suggesting different role for these cells in hematopoiesis. The divergences between CD133+ and CD34+ cell populations may be more obvious in those cells from BM or PB. It may also be that the divergences between CD133+ and CD34+ cells reflect specific primitiveness of CB CD133+ cells.

**CD133+ and CD34+ cells demonstrate similar total colony forming potential (studies I and II)**

CFU assay was used to identify primitive hematopoietic cells from CD133+, CD34+, CD133-, CD34- and mononuclear cells. A similar number of total CFUs were formed by CD133+ and CD34+ cells. CD133+ cells formed almost only multipotent granulocyte-erythroid-macrophage-megakaryocyte (38%) and granulocyte-macrophage colonies (58%). Burst-forming colonies represented 4.2% of the progeny in CD133+ cells, whereas no erythroid colonies were observed. A vast majority of CD34+ cell progeny was also multipotential, 33% granulocyte-erythroid-macrophage-megakaryocyte colonies and 58% granulocyte-macrophage colonies. A little more of both the burst-forming (7.2%) and unipotent erythroid colonies (1.4%) were formed by CD34+ cells (Figure 8). Combining the CFU assay and gene expression profiling results it seems likely that CD133+ cells are more primitive in their nature when compared to CD34+ cells.
RESULTS AND DISCUSSION

Reanalysis of the original gene expression data focuses on the similarities between CD133+ and CD34+ cells

Due to the fast development of gene expression analysis and data normalization methods, the original data from microarray analysis (study I and II) were reanalyzed using a more modern method. In the reanalysis, the top 500 genes with differential expression in CD133+/CD133- or CD34+/CD34- cells were chosen. After discharging putative false positive results, based on the q-value, 492 genes in the CD133+ gene expression profile and 425 genes in the CD34+ gene expression profile were considered valid (Figure 9). Of these, CD133+/CD133- and CD34+/CD34- cells had 283 differentially expressed genes in common. In comparison to CD133- and CD34- data, 118 genes were overexpressed and 165 genes were underexpressed in CD133+ and CD34+ cells.

Reanalysis of the microarray data verified the overexpression of 67 genes in CD133+ and CD34+ cells when compared to CD133- and CD34- cells. These include KIT, FLT3, MEIS1, HOX9A, GATA2 and SOCS2 genes with a significant role in HSC regulation. Similarly, 73 genes downregulated in CD133+ and CD34+ cells, when compared to CD133- and CD34- cells, were the same between the original and reanalyzed data. No genes that were defined as overexpressed in CD133+ and CD34+ cells by the original data analysis were shown to be underexpressed after reanalysis, or vice versa. However, the reanalysis of the data identified 51 new genes with a higher expression level in CD133+ and CD34+ cells than in

Figure 8. Clonogenic progenitor capacity of CD34+ and CD133+ cells by CFU assay. Abbreviations: BFU, burst-forming erythroid; CFU, colony-forming unit; GEMM, granulocyte-erythroid-macrophage-megakaryocyte; GM, granulocyte-macrophage.
CD133+ and CD34+ cells (data not shown). Moreover, 92 new genes were downregulated in CD133+ and CD34+ cells when compared to CD133− and CD34− cells. The role of these novel genes in the regulation of HSCs requires further investigation.

The association with a biological process was defined for the set of differentially expressed genes in the original data and the reanalyzed data. Only the genes similarly expressed in CD133+ and CD34+ cells were considered. Annotation was performed anew to both data sets (Figure 10). In both cases, the biological processes representative of CD133+ and CD34+ cells were clearly different from their negative counterparts. The associated biological processes enriched in CD133+ and CD34+ cells have been shown to be characteristics to HSC populations106,109,199. Many of the related biological processes were the same between the original data and the reanalyzed data, especially in the processes that represent CD133− and CD34− cells.

The causes for genes with higher expression level in CD133+ and CD34+ cells than in CD133− and CD34− cells related to catalytic activity (38 genes), cytokine activity (5 genes) and endopeptidase/protease inhibitor activity (4 genes). Cytokine activity-associated genes SOCS2, TNFSF4, TRIP6, PXDN and IPO11 were the same as identified in the original data (studies I and II). The 4 protease inhibitor genes were SERPING1, SERPINE2, SPINK2 and TFPI. Similarly to other protease inhibitors200,201,
Figure 10. Biological processes represented by differentially expressed genes in CD133<sup>+</sup> and CD34<sup>+</sup> cells when compared to CD133<sup>-</sup> and CD34<sup>-</sup> cells. A) Original data (study II) and B) reanalyzed data.

these genes may have a role in the modulation of cytokine expression. Their overexpression in HSCs has been previously shown<sup>112,115,188,193</sup>. Forty-nine of the genes overexpressed in CD133<sup>+</sup> and CD34<sup>+</sup> cells were associated with intracellular compartments, especially to cytoplasm (25 genes) (Table 4).

The genes downregulated in CD133<sup>+</sup> and CD34<sup>+</sup> cells when compared to CD133<sup>-</sup> and CD34<sup>-</sup> cells were associated with cell communication (32 genes), response to other organism (17 genes), apoptosis (16 genes) and cell activation (7 genes). Moreover, the encoded protein products were associated with protein binding (42 genes) and receptor activity (20 genes). The most strikingly emphasized cellular compartment for the encoded gene products was cell membrane (55 genes). CD133<sup>-</sup> and CD34<sup>-</sup> cell populations include lymphocytes, mononuclear phagocytes, dendritic cells and granulocytes that act in the immune system. It is crucial for these cells to have active signaling and communication with both other cells and the matrix. The immune response is dependent on detected changes in the environment and the activation of the cell is initiated by extrinsic factors.
Table 4. The most common cellular compartments for products encoded by expressed genes in CD133\(^+\) and CD34\(^+\) cells.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Name</th>
<th>Hs.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cytoplasm</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALDH1A1</td>
<td>Aldehyde dehydrogenase 1 family, member A1</td>
<td>76392</td>
</tr>
<tr>
<td>ANT</td>
<td>Aminotransferase</td>
<td>102</td>
</tr>
<tr>
<td>ALG8</td>
<td>Asparagine-linked glycosylation B homolog</td>
<td>503368</td>
</tr>
<tr>
<td>BCL11A</td>
<td>B-cell CLL/lymphoma 11A (zinc finger protein)</td>
<td>370549</td>
</tr>
<tr>
<td>BAALC</td>
<td>Brain and acute leukemia, cytoplasmic</td>
<td>533446</td>
</tr>
<tr>
<td>BCA1</td>
<td>Branched chain aminotransferase 1, cytosolic</td>
<td>438993</td>
</tr>
<tr>
<td>CLGN</td>
<td>Calmodulin</td>
<td>86368</td>
</tr>
<tr>
<td>CHST13</td>
<td>Carbohydrate (chondroitin 4) sulfotransferase 13</td>
<td>292375</td>
</tr>
<tr>
<td>CPA3</td>
<td>Carboxypeptidase A3 (mast cell)</td>
<td>646</td>
</tr>
<tr>
<td>CTHRC1</td>
<td>Collagen triple helix repeat containing 1</td>
<td>405614</td>
</tr>
<tr>
<td>GGH</td>
<td>Gamma-glutamyl hydrolase (conjugase, fattyacylglycerol hydrolase)</td>
<td>78619</td>
</tr>
<tr>
<td>IPO11</td>
<td>Importin 11</td>
<td>482269</td>
</tr>
<tr>
<td>JUP</td>
<td>Junction plakoglobin</td>
<td>51474</td>
</tr>
<tr>
<td>MF3</td>
<td>Malic enzyme 3, NADP(+)-dependent, mitochondrial</td>
<td>199743</td>
</tr>
<tr>
<td>PHYH</td>
<td>Phytanoyl-CoA 2-hydroxylase</td>
<td>498732</td>
</tr>
<tr>
<td>PKD2</td>
<td>Polycystic kidney disease 2 (autosomal dominant)</td>
<td>181272</td>
</tr>
<tr>
<td>PGDS</td>
<td>Prostaglandin D2 synthase, hematopoietic</td>
<td>128433</td>
</tr>
<tr>
<td>RAB34</td>
<td>RAB34, member RAS oncogene family</td>
<td>301853</td>
</tr>
<tr>
<td>AHCY</td>
<td>S-adenosylhomocysteine hydrolase</td>
<td>388004</td>
</tr>
<tr>
<td>SLC27A2</td>
<td>Solute carrier family 27 (fatty acid transporter), member 2</td>
<td>11729</td>
</tr>
<tr>
<td>SOCS2</td>
<td>Suppressor of cytokine signaling 2</td>
<td>485572</td>
</tr>
<tr>
<td>GALNT7</td>
<td>UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 7</td>
<td>127407</td>
</tr>
<tr>
<td><strong>Intracellular</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NARS2</td>
<td>Asparagine-IRNA synthetase 2 (mitochondrial)(putative)</td>
<td>503389</td>
</tr>
<tr>
<td>CREB3L4</td>
<td>CAMP responsive element binding protein 3-like 4</td>
<td>372924</td>
</tr>
<tr>
<td>CDC47</td>
<td>Cell division cycle associated 7</td>
<td>470654</td>
</tr>
<tr>
<td></td>
<td>Transcribed locus</td>
<td>47312</td>
</tr>
<tr>
<td>CCNB1IP1</td>
<td>Cyclin B1 interacting protein 1</td>
<td>107003</td>
</tr>
<tr>
<td>GATA2</td>
<td>GATA binding protein 2</td>
<td>367725</td>
</tr>
<tr>
<td>GINS2</td>
<td>GINS complex subunit 2 (Psf2 homolog)</td>
<td>433180</td>
</tr>
<tr>
<td>HLF</td>
<td>Hepatic leukemia factor</td>
<td>156952</td>
</tr>
<tr>
<td>HOXA9</td>
<td>Homeobox A9</td>
<td>110637</td>
</tr>
<tr>
<td>MEIS1</td>
<td>Meis1, myeloid ecotropic viral integration site 1 homolog</td>
<td>526754</td>
</tr>
<tr>
<td>MRPS27</td>
<td>Mitochondrial ribosomal protein S27</td>
<td>482491</td>
</tr>
<tr>
<td>NDN</td>
<td>Necdin homolog</td>
<td>50130</td>
</tr>
<tr>
<td>NRP1</td>
<td>Nuclear receptor Interacting protein 1</td>
<td>155017</td>
</tr>
<tr>
<td>NAPIL3</td>
<td>Nucleosome assembly protein 1-like 3</td>
<td>21365</td>
</tr>
<tr>
<td>NUDT11</td>
<td>Nudix (nucleoside diphosphate linked moiety X)-type motif 11</td>
<td>200016</td>
</tr>
<tr>
<td>PLS3</td>
<td>Plastin 3 (T Isoform)</td>
<td>496622</td>
</tr>
<tr>
<td>PLAG1</td>
<td>Pleiomorphic adenoma gene-like 1</td>
<td>444975</td>
</tr>
<tr>
<td>PRIM1</td>
<td>Primase, polyplase 1, 49kDa</td>
<td>534339</td>
</tr>
<tr>
<td>RPP40</td>
<td>Ribonuclease P 40kDa subunit</td>
<td>511756</td>
</tr>
<tr>
<td>SCR1</td>
<td>Secernin 1</td>
<td>520740</td>
</tr>
<tr>
<td>SH3BP4</td>
<td>SH3-domain binding protein 4</td>
<td>516777</td>
</tr>
<tr>
<td>SMAD1</td>
<td>SMAD family member 1</td>
<td>519005</td>
</tr>
<tr>
<td></td>
<td>Transcribed locus</td>
<td>492445</td>
</tr>
<tr>
<td>MYB</td>
<td>V-myb myeloblastosis viral oncogene homolog</td>
<td>531941</td>
</tr>
<tr>
<td>ZNRF1</td>
<td>Zinc and ring finger 1</td>
<td>427284</td>
</tr>
<tr>
<td>ZNF521</td>
<td>Zinc finger protein 521</td>
<td>116935</td>
</tr>
<tr>
<td>ZNF711</td>
<td>Zinc finger protein 71</td>
<td>326801</td>
</tr>
</tbody>
</table>
The main functional categories related to the CD133+ and CD34+ gene expression profiles were the same in the original data and the reanalyzed data using the most recent annotation. There are several reasons to explain the differences between the original data and the reanalyzed data. These are: the difference between the redesigned probe sets and original Affymetrix probe sets, different preanalysis and normalization of the data, and the criteria of the differentially expressed genes. Redesigned probe sets or sequence matched probes have shown a 30-50% discrepancy of results in differential gene expression analysis when compared to analysis using the original Affymetrix probe sets.\footnote{181,202} It is important to notice that the differences between the original and reanalyzed data sets of differentially expressed genes in CD133+/CD133- and CD34+/CD34- cells are not necessarily faulty results. Well-characterized genes expressed in CD133+/CD34+ cells are not all among the list of the most highly expressed genes in the reanalysis due to the selection criteria that was used. Thus, the differential gene expression that was shown in the original data, but not in reanalyzed data, is still in the main part usable. However, redesigned probes have the improved capacity to detect the expressed genes. Thus, the reanalysis of the gene expression data may help in identification of new genes and provide novel data for further studies.

**CD133+ and CD34+ cells have divergent roles in hematopoietic regeneration (studies I and II)**

The divergent gene expression between CD133+ and CD34+ cells was revealed by direct pairwise analysis. The represented analysis was specially focused on genes associated with hematopoiesis, cell cycle and cell adhesion. CD133+ cells had a higher expression level of CD133, CD34, CD135 (FLT3) and CD117 (KIT) genes. The higher expression of CD34 in the CD133+ cell population may be explained by the higher abundance of CD34 on the cell surface of CD133+ cells. Nearly all CD133+ cells are CD34-positive, whereas the CD34+ cell population includes lineage committed progenitors that carry a lower number of CD34 molecules on their cell surface.

The hematopoietic lineage-specific markers were downregulated in CD133+ and CD34+ cells when compared to their negative counterparts. However, there was a difference in the level of underexpression between CD133+ and CD34+ cells (Table 5). CD133+ cells expressed IL3R (CD123), a marker typical for common myeloid progenitors, pronormoblasts and
megakaryoblasts. The other genes encoding megakaryoblast markers, such as *IL-11R* or *CD126*, were not expressed in CD133+ cells. On the other hand, one or more of the CD34+ sample replicates expressed several genes related to hematopoietic differentiation. The complete list of genes related to hematopoiesis is shown in Table 5.

The lineage differentiation marker expression may be connected to the cells in the CD133+ and CD34+ subpopulations. Hence, these results suggest that the CD34+ cell population contains a higher proportion of pro T cells, other differentiating T cells, promonoblasts, monoblasts and megakaryocytes than the CD133+ cell population (Figure 1). On the other hand, it has been suggested that HSCs express lineage differentiation marker genes at a low level to enable the fast differentiation of HSCs when a differentiated cell type is needed in the blood system. It may also be that CD34+ cell population consists of a higher number of short-term HSCs with better readiness to form all types of blood cells.

CD34+ cells expressed a wider variety of genes encoding cell adhesion molecules, such as selectin E and selectin P. Selectins are important for cell trafficking. CD34+ cells may thus have greater capacity to be mobilized. However, the engraftment potential of CB HSCs has been delayed when compared to HSCs from other sources. The presence of very late antigen 4 has been proven crucial for the homing of HSCs. The expression of gene coding for very late antigen 4 was higher in CD133+ cells.

The gene expression profile of CD133+ cells revealed the expression of many genes related to the cell cycle (study I). CD133+ cells expressed 20 cell cycle genes on a higher level than CD34+ cells. These products of the cell cycle genes have inhibitor as well as promoter effects. The results suggest that both quiescent cells and actively cycling cells are found in the CD133+ cell population. Long-term HSCs are mostly considered to be noncycling, but that is not as critical a parameter in CB HSCs. Hence, CB HSC in an active cell cycle state can have self-renewing capacity.

CD133+ and CD34+ cells had a high number of commonly expressed genes typical of HSCs. They also had many overlapping biological processes. Both CD133+ and CD34+ cells have been successfully used in the reconstruction of the blood system in irradiated patients’ marrow. Thus both cell populations contain cells with long-term repopulation and short-term repopulation potential. The direct comparison of CD133+ and CD34+ cells revealed very delicate differences between the cell types. Previous studies have shown that differences in the gene expression
RESULTS AND DISCUSSION

Profile reflect true changes between closely related cell populations\textsuperscript{205,206}. Thus, the differences demonstrated in expression profiles of CD133\textsuperscript{+} and CD34\textsuperscript{+} cells suggest functional divergences between the cell types.

The gene expression profiles indicate that faster repopulation of the recipient could be obtained after CD34\textsuperscript{+} cell transplantation. Due to the more primitive nature of CD133\textsuperscript{+} cells, selected CD133\textsuperscript{+} cell population may not include enough cells readily available for development of all blood cell types or they may have poorer capacity to travel into the BM,

\textbf{Table 5.} Differential expression of hematopoietic cell marker genes in CD133\textsuperscript{+} and CD34\textsuperscript{+} cells.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Name</th>
<th>Unigene cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher expression level in CD133\textsuperscript{+} cells</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD10/MME</td>
<td>Membrane metallo-endopeptidase</td>
<td>Hs.307734</td>
</tr>
<tr>
<td>CD34</td>
<td>CD34 molecule</td>
<td>Hs.374990</td>
</tr>
<tr>
<td>CD49d/ITGA4</td>
<td>Integrin, alpha 4 (alpha 4 subunit of VLA-4 receptor)</td>
<td>Hs.694732</td>
</tr>
<tr>
<td>CD49e/ITGA5</td>
<td>Integrin, alpha 5 (fibronectin receptor, alpha polypeptide)</td>
<td>Hs.505654</td>
</tr>
<tr>
<td>CD59</td>
<td>CD59 molecule, complement regulatory protein</td>
<td>Hs.653685</td>
</tr>
<tr>
<td>CD117/KIT</td>
<td>V-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog</td>
<td>Hs.479754</td>
</tr>
<tr>
<td>CD135/FLT3</td>
<td>Fms-related tyrosine kinase 3</td>
<td>Hs.507590</td>
</tr>
<tr>
<td>Higher expression level in CD34\textsuperscript{+} cells</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD1A</td>
<td>CD1a molecule</td>
<td>Hs.1309</td>
</tr>
<tr>
<td>CD1B</td>
<td>CD1b molecule</td>
<td>Hs.1310</td>
</tr>
<tr>
<td>CD1D</td>
<td>CD1d molecule</td>
<td>Hs.1799</td>
</tr>
<tr>
<td>CD2</td>
<td>CD2 molecule</td>
<td>Hs.523500</td>
</tr>
<tr>
<td>CD3E</td>
<td>CD3e molecule, epsilon</td>
<td>Hs.3003</td>
</tr>
<tr>
<td>CD3G</td>
<td>CD3g molecule, gamma</td>
<td>Hs.2259</td>
</tr>
<tr>
<td>CD8A</td>
<td>CD8a molecule</td>
<td>Hs.85258</td>
</tr>
<tr>
<td>CD8B</td>
<td>CD8b molecule</td>
<td>Hs.405667</td>
</tr>
<tr>
<td>CD9</td>
<td>CD9 molecule</td>
<td>Hs.114286</td>
</tr>
<tr>
<td>CD11b/ITGAM</td>
<td>Integrin, alpha M (complement component 3 receptor 3 subunit)</td>
<td>Hs.172631</td>
</tr>
<tr>
<td>CD14</td>
<td>CD14 molecule</td>
<td>Hs.163867</td>
</tr>
<tr>
<td>CD24</td>
<td>CD24 molecule</td>
<td>Hs.694721</td>
</tr>
<tr>
<td>CD38</td>
<td>CD38 molecule</td>
<td>Hs.479214</td>
</tr>
<tr>
<td>CD41/ITGA2B</td>
<td>Integrin, alpha 2b (platelet glycoprotein IIb)</td>
<td>Hs.411312</td>
</tr>
<tr>
<td>CD42a/GP9</td>
<td>Platelet glycoprotein IX</td>
<td>Hs.1144</td>
</tr>
<tr>
<td>CD42d/GP5</td>
<td>Platelet glycoprotein V</td>
<td>Hs.73734</td>
</tr>
<tr>
<td>CD49/ITGA6</td>
<td>Integrin, alpha 6</td>
<td>Hs.133397</td>
</tr>
<tr>
<td>CD61/ITGB3</td>
<td>Integrin, beta 3 (platelet glycoprotein IIIa, antigen CD61)</td>
<td>Hs.218040</td>
</tr>
<tr>
<td>CD64/FCGR1A</td>
<td>Fc fragment of IgG, high affinity 1a, receptor</td>
<td>Hs.77424</td>
</tr>
<tr>
<td>CD115/CSF1R</td>
<td>Colony stimulating factor 1 receptor</td>
<td>Hs.654394</td>
</tr>
<tr>
<td>CD116/CSF2RA</td>
<td>Colony stimulating factor 2 receptor, alpha</td>
<td>Hs.520937</td>
</tr>
<tr>
<td>CD121b/IL1R2</td>
<td>Interleukin 1 receptor, type II</td>
<td>Hs.25333</td>
</tr>
<tr>
<td>CD127/IL7R</td>
<td>Interleukin 7 receptor</td>
<td>Hs.693675</td>
</tr>
<tr>
<td>CSF1</td>
<td>Colony stimulating factor 1 (macrophage)</td>
<td>Hs.591402</td>
</tr>
<tr>
<td>CSF2</td>
<td>Colony stimulating factor 2 (granulocyte-macrophage)</td>
<td>Hs.1349</td>
</tr>
<tr>
<td>IL1A</td>
<td>Interleukin 1, alpha</td>
<td>Hs.1722</td>
</tr>
<tr>
<td>TPO</td>
<td>Thyroid peroxidase</td>
<td>Hs.467554</td>
</tr>
<tr>
<td>TNF</td>
<td>Tumor necrosis factor</td>
<td>Hs.241570</td>
</tr>
</tbody>
</table>
or both. CD133\(^+\) cells, as a more primitive source of HSCs, could be better suited for HSC expansion and gene therapy. Due to the limited number of transplantations performed using selected CD133\(^+\) and CD34\(^+\) cells, it is not possible to say whether there is a difference in the repopulation after transplantation. It should be kept in mind that despite the divergent biological capacities, binding of an antibody to the CD133 or CD34 molecules may influence cell function.

**N-glycosylation-related gene expression is similar in CD133\(^+\) and CD34\(^+\) cells and correlates with the N-glycome profile (study III)**

CD133\(^+\) and CD34\(^+\) cells had characteristic N-glycosylation-related gene expression when compared to CD133\(^-\)/CD34\(^-\) cells (study III). Moreover, mass spectrometric analysis of CD133\(^+\) and CD133\(^-\) cells, combined with results from NMR and exoglycosidase digestion, revealed differences in the glycosylation of these cells. CD133\(^+\) cells contained significantly more fully processed, especially biantennary type N-glycans than CD133\(^-\) cells (Figure 11). Yet, fully processed multiantennary N-glycans were fewer in CD133\(^+\) cells than in CD133\(^-\) cells. The most frequent glycosylation of CD133\(^+\) cells was unprocessed, high-mannose type N-glycans. High mannose type N-glycans were more abundant in CD133\(^+\) cells than in CD133\(^-\) cells.

Enrichment of high mannose type N-glycans in CD133\(^+\) cells was evident in the comparison of CD133\(^+\) cells to CD8\(^+\) cells which represent a mature leukocyte population (data not shown). However, more analyses are needed to confirm that the differences in the relative abundance of neutral N-glycans are due to tissue-specific glycosylation changes.

**Characteristic mannosylation pattern of CD133\(^+\) and CD34\(^+\) cells (study III)**

The differential expression of one mannosidase and three glycosyltransferase encoding genes were related to the changes observed in the CD133\(^+\) and CD133\(^-\) cell glycome. The differential gene expression was observed in both CD133\(^+\) and CD34\(^+\) cells as compared to their negative counterparts. CD133\(^+\) and CD34\(^+\) cells, but not CD133\(^-\) and CD34\(^-\) cells, lacked the expression of MAN1C1 encoding one of the key \(\alpha1,2\)-mannosidases participating in N-glycan maturation. No difference was seen in the expression of other \(\alpha1,2\)-mannosidase encoding genes
**MAN1A1, MAN1A2 and MAN1B1**, yet their expression was detected in CD133\(^+\) and CD34\(^+\) cells as well as CD133\(^-\) and CD34\(^-\) cells.

MAN1 enzymes remove \(\alpha_1,2\)-mannosidases from different positions in N-glycans, and their functions are highly overlapping (Figure 6). MAN1A, MAN1B and MAN1C are located in different genes on chromosomes 6 and 1, and they have independent regulatory pathways. The unique specificity of MAN1C1 encoded \(\alpha_1,2\)-mannosidase 1C is that it prefers the removal of terminal \(\alpha_1,2\)-linked mannose on the \(\alpha_1,3\) branch in Man\(_9\)GlcNAC\(_2\) and Man\(_9\)GlcNAC\(_2\) isomer B\(^{207}\). The \(\alpha_1,2\)-mannosidases 1A and 1B remove terminal \(\alpha_1,2\)-linked mannose from the \(\alpha_1,6\) branch from Man\(_9\)GlcNAC\(_2\) isomer B before the \(\alpha_1,3\) branch is demannosylated. Hence, the absence of MAN1C1 expression in CD133\(^+\) and CD34\(^+\) cells may affect the enrichment of high mannose N-glycans. It is likely that most of the high mannose N-glycans seen in the CD133\(^+\) and CD133\(^-\) cell glycan profile were intracellular. N-glycan profiles include all N-glycans, not only those on the cell surface. However, labeling with mannose binding lectins *Pisum sativum* agglutinin and *Hippeastrum hybrid* agglutinin confirmed the presence of mannose-rich N-glycans on the cell surface.

High mannose N-glycans promote proper protein folding\(^{208}\). The high mannose N-glycans transported to the cell surface are known to be demannosylated. A half-life of 6.7 hours for cell surface high mannose

![Figure 11](image-url)  
*Figure 11.* Schematic representation of the differences in CD133\(^+\) and CD133\(^-\) cell N-glycan profiles. In CD133\(^+\) cells, high mannose N-glycans and biantennary complex type N-glycans were the most enriched N-glycan structures. In contrast, the proportion of low mannose type N-glycans and multiantennary structures greater than biantennary complex N-glycans was increased in CD133\(^-\) cells. Monosaccharide symbols: ■ = N-acetyl-D-glucosamine; ○ = D-mannose. Glycosidic linkages are indicated by lines connecting the monosaccharides.
N-glycans has been observed in the hepatic cell line\textsuperscript{209}. On the other hand, 4TO7 tumor cells are capable of spreading to the lungs and they have been shown to have more high mannose and N-glycans on their cell surface than nonmetastatic tumor cells\textsuperscript{210}. Further, it has been shown that sialic acids have no effect on the migration and invasion properties of 4TO7 cells. Thus, high mannose N-glycans may have yet unknown functions.

The 5\% increase of high mannose type N-glycans demonstrated in CD133\(^+\) cells when compared to CD133\(^-\) cells may correspond to a variety of glycoproteins. There may be a greater number of recently transported proteins on the surface of CD133\(^+\) cells. Hence, there may simply be more time for trimming mannose residues to form the low mannose type N-glycans more abundant in CD133\(^-\) cells. If there is a greater proportion of mannose-rich glycans on the surface of CD133\(^+\) cells, it could be due to the lower number of mannose trimming enzymes in these cells. The unique mannosylation of CD133\(^+\) cells may be useful for classification of a primitive cell type. The lack of MAN1C1 expression may function as a marker for HSCs.

**Differential expression of N-acetylglucosaminyl-transferase genes and the enrichment of biantennary complex type N-glycans in CD133\(^+\) cells (study III)**

Differential expression of two genes encoding N-glycan backbone branching N-acetylglucosaminyltransferase enzymes was demonstrated between CD133\(^+\)/CD133\(^-\) and CD34\(^+\)/CD34\(^-\) cells. \textit{MGAT2} was overexpressed 1.9-fold in CD133\(^+\) cells when compared to CD133\(^-\) cells. \textit{MGAT2} overexpression was 1.7-fold in CD34\(^+\) cells when compared to CD34\(^-\) cells. \textit{MGAT4A} was underexpressed 2.8-fold in CD133\(^+\) cells and 2.0-fold in CD34\(^+\) cells when compared to their negative counterparts. The increase in the amount of biantennary complex type N-glycans demonstrated in CD133\(^+\) cells correlated with the higher expression of \textit{MGAT2}. Furthermore, a lower number of triantennary and other multiantennary N-glycans were seen in CD133\(^+\) cells. \textit{MGAT4} gene encoding N-acetylglucosaminyltransferase 4 participates in the biosynthesis of triantennary and other multiantennary complex type N-glycans. Thus, underexpression of \textit{MGAT4A} may be related to the reduction of multiantennary N-glycans in CD133\(^+\) and CD34\(^+\) cells. Complex type N-glycans seem to be developmentally important as \textit{MGAT2}-null mice suffer from hematological and osteogenic abnormalities and die soon after birth. This
indicates that complex type N-glycans could have a function in the interaction between HSCs and BM osteoblasts. High binding of *Phaseolus vulgaris*-derived Phytohemagglutinin-L lectin, specific to large complex type N-glycans, demonstrated the presence of large complex type N-glycans on leukocytes including the CD34+ cells.

**Despite the differential expression of genes coding for β1,4-galactosyltransferases there are no differences in the β1,4-galactosylation of CD133+ and CD133- cells (study III)**

Type 2 N-acetyllactosamine dominated as glycan backbone structure in CD133+ and CD133- cells. Mass spectrometric analysis results were verified by binding of Ricinus communis agglutinin-1 lectin. Moreover, type 2 N-acetyllactosamine were β1,4-galactosylated as β1,3-linked galactose was not detected. Genes encoding β1,4-galactosyltransferases such as *B4GALT1*, *B4GALT3* and *B4GALT4* were expressed in CD133+, CD133-, CD34+ and CD34- cells. *B4GALT3* was downregulated in CD133+ and CD34+ cells by 2.3- and 2.0-fold respectively, when compared to their negative counterparts.

The differential expression of β1,4-galactosyltransferase encoding genes was not shown to have any effect on N-glycosylation. Yet, it may not be ruled out that the differential expression of β1,4-galactosyltransferase encoding genes still have a significant function in the galactosylation of specific proteins not detected in glycan profiling, or have an impact on the galactosylation of other glycan types such as O-glycans or glycolipids.

**CD133+ and CD34+ cell specific biosynthesis of N-glycan terminal epitopes (study III)**

The complex type N-glycans, biantennary and larger, were very often α2,6-sialylated in CD133+ and CD133- cells. *ST6Gal1* encoding α2,6-sialyltransferase was similarly expressed in CD133+/CD133- cells, and CD34+/CD34- cells. Lectin *Sambucus nigra* agglutinin recognizing α2,6-linked sialic acid bound to 98% of the leukocytes including CD34+ cells. Thus, α2,6-sialylation was a dominant terminal epitope on CD133+ and CD133- cells. Moreover, α2,3-sialylation had increased on CD133+ cells when compared to CD133- cells. The α2,3-sialidase digestion disclosed the presence of some solely α2,3-sialylated N-glycans in CD133+ cells.
The $\alpha_{2,3}$-sialidase treatment was not able to entirely desialylate N-glycans in CD133$^-$ cells. CD133$^+$ and CD34$^+$ cells overexpressed the $\alpha_{2,3}$-sialyltransferase encoded by $ST3GAL6$ when compared to their negative counterparts (CD133$^-$ and CD34$^-$ cells). $ST3GAL6$ expression was 3.9-fold higher in CD133$^+$ cells than CD133$^-$ cells, and 1.6-fold higher in CD34$^+$ cells than CD34$^-$ cells. Lectin $Maackia amurensis$ agglutinin with $\alpha_{2,3}$-linked sialic acid specificity bound to 62% of the leukocytes and CD34$^+$ cells demonstrating the presence of $\alpha_{2,3}$-sialyl-N-acetyllactosamine structures on the cell surface.

$\alpha_{2,3}$- and $\alpha_{2,6}$-sialyltransferases compete for the same N-glycan substrates. Lowered amount of $\alpha_{2,6}$-sialyltransferase and decreased $\alpha_{2,6}$-sialylation in murine T cells was suggested to be caused by substrate competition with $\alpha_{1,3}$-galactosyltransferase$^{211}$. However, $\alpha_{1,3}$-galactosyltransferase is not present in human cells and similar substrate competition is not relevant in humans. The lower relative abundance of $\alpha_{2,6}$-sialylation in CD133$^+$ cells is likely to be caused by increased $\alpha_{2,3}$-sialylation. $ST6Gal1$ or $ST3Gal6$ have not been shown crucial for development. B cells express CD22 (Siglec2) that specifically binds to $\alpha_{2,6}$-linked sialic acid. However, Siclec2 functions only in activated B cells, while it is masked by the cell's own $\alpha_{2,6}$-sialylation in resting B cells$^{212}$. B cells that have reduced sialylation have been shown more susceptible to apoptosis$^{213}$. Similar masking by sialic acids could also exist in other cell types to prevent Fas receptor ligand driven apoptosis. HSCs upregulate Fas upon active cell cycle, but remain resistant to Fas-induced suppression$^{214}$. Fas ligand coating of HSCs improves their engraftment capacity by two-fold$^{215}$. In addition, HSCs have shown to secrete Fas ligand which targets the marrow cells of the recipient in transplantation$^{215}$.

Sialylation has been suggested to play a role in the initial attachment of HSCs to BM osteoblasts$^{158}$. However, the participation of any single adhesion molecule in the homing of HSCs has not been unequivocally shown$^{216}$. Glycosylation-based recognition may be involved in the initial steps of homing, but the binding of specific proteins are required for engraftment. The interaction may involve expression of stem cell factor and stromal cell-derived factor 1, activation of lymphocyte function-associated antigen 1, very late antigen 4, very late antigen 5 and CD44, cytoskeleton rearrangements and secretion of membrane type 1–matrix metalloproteinases$^{217}$. In the future, it would be interesting to find out whether $\alpha_{2,3}$-sialylation is the only form of sialylation in certain proteins. It is also not yet known if solely $\alpha_{2,3}$-sialylated structures are typical of BM CD133$^+$ cells as well. More studies are needed to reveal whether
the differential sialylation observed in CB CD133+ cells is beneficial

to cell function or if these cells would benefit from a higher level of

sialylation.

Only a small amount of CD34+ cells have been shown to bind to

hyaluronan in BM. The α2,3-sialylation of N-glycans has been shown

reduce the binding of CD44 molecule to hyaluronan in CHO cell line.218

Enriched α2,3-sialylation may have a role in the mobilization of HSCs
to the circulatory system. CD34+ cells carry a specific glycoform of

CD44 molecule on their cell surface. It has been shown to have over
5-fold L-selectin binding when compared to P-selectin glycoprotein

ligand 1.168

Selectin ligand sialyl-Lewisx is dependent on α2,3-sialylation and α1,3-

fucosylation. In addition to common α1,6-fucosylation of N-glycan

core structures, one or two more fucose residues were revealed in

CD133+ and CD133- cell N-glycans by mass spectrometric analysis. The

fucose was expected to be α1,2- or α1,3-linked to the N-glycans as no

type 1 N-acetyllactosamine was found on CD133+ or CD133- cells. FUT1

and FUT2 both encode α1,2-fucosyltransferase, but their expression

was not reliably analyzed. FUT1 expression was not detected in CD133+/CD34+
cells or their negative counterparts. For FUT2 there was no probe
available on the array. Ulex europaeus agglutinin I lectin with speci-

city towards α1,2-linked fucose residues labeled 53% of the leukocytes.

Impaired α1,3-fucosylation and decreased α1,3-fucosyltransferase

expression has been shown to be characteristic of CB CD34+ cells.163,164

In order to define α1,3-linked fucose on CD133+ or CD133- cell N-glycans,

leukocytes were labeled with Lotus tetragonolobus agglutinin lectin.

It bound to only 6% of leukocytes. Moreover only FUT4, but no FUT7

expression was detected in CD133+, CD34+, CD133- and CD34- cells.

FUT4 was expressed similarly in all cell types. FUT7 encodes the main

α1,3-fucosyltransferase responsible for the synthesis of sialyl-Lewisx,

but FUT4 protein product is also able to synthesize it. These results
suggest that poor α1,3-fucosylation is typical for the CB leukocytes
including CD34+ cells.
CONCLUSIONS

The positive selection of HSCs has provided an attractive alternative to T cell depletion of transplantable cells. The selection of CD34\(^+\) cells can be used in this process. However, the selection of CD133\(^+\) cells has also shown to be promising in the enrichment of HSCs\(^{220-222}\). Although, CD133\(^+\) and CD34\(^+\) cells are highly similar cell populations there may be differences in their genetic regulation. In previous studies, HSCs from BM, PB and CB have shown differential gene expression related to divergent biological properties operative in HSCs\(^{10,155,223}\). Increased knowledge of HSC biology is crucial to improve stem cell-based therapies.

The present studies characterize human CB-derived CD133\(^+\) and CD34\(^+\) cells on a genomic level and provide gene expression profiles of CD133\(^+\) and CD34\(^+\) cells when compared to CD133\(^-\) and CD34\(^-\) cells, respectively. The integrated CD133\(^+/\)CD34\(^+\) cell gene expression profiles reveal novel genes to specify HSCs. Moreover, the transcripts that are expressed differentially between CD133\(^+\) and CD34\(^+\) cells are associated with cell cycle, cell adhesion and differentiation. This finding suggests that the divergent gene expression patterns may be adjusting CD133\(^+\) and CD34\(^+\) cells in different ways in order to encompass hematopoiesis. CD133\(^+\) cells seem to be more primitive in their nature and possibly have a wider differentiation potential than CD34\(^+\) cells. CD133\(^+/\)CD34\(^+\) cells express numerous transcripts coding for putative transmembrane proteins that may serve as markers for HSCs. However, during this study, specific markers to identify HSCs could not be found on the cell surface of CD133\(^+\) or CD34\(^+\) cells.

These studies also demonstrate N-glycan structure profile typical for CD133\(^+\) cells. CD133\(^+\) cells have enriched amount of high mannose type and biantennary complex type N-glycans when compared to CD133\(^-\) cells. CD133\(^+\) cells also have solely \(\alpha_2,3\)-sialylated N-glycans. Comparison of N-glycan structure and gene expression profiles identify key genes regulating the CD133\(^+\) cell-specific N-glycan synthesis. The expression pattern of genes encoding glycosyltransferases and glucosidases is similar in CD133\(^+\) and CD34\(^+\) cells, suggesting a similar N-glycosylation in CD133\(^+\) and CD34\(^+\) cells. Furthermore, the specific binding properties of various lectins are utilized to reveal the glycan determinants on the cell surface.
Further studies to identify proteins that carry the specific N-glycan structures in CD133+ cells could prove useful in the recognition and isolation of HSCs. Moreover, the role of glycans and glycan binding proteins in the communication between HSCs and their environment should be elucidated. The modification of HSC glycosylation could increase HSC homing and engraftment or target them to specific tissues. In conclusion, these results provide new knowledge of CD133+ and CD34+ cells that could help to design cells with enhanced biological properties and to further the use of HSCs in novel therapeutic applications.
ACKNOWLEDGEMENTS

This work was carried out at the Finnish Red Cross Blood Service, Research and Development, in Helsinki during the years 2004-2007. Partial funding support was provided by the Finnish Funding Agency for Technology and Innovation.

I am grateful to the Director of the Institute, Docent Jukka Rautonen, and Department Head, Kari Aranko, for providing excellent research facilities. I devote my warmest thanks to my supervisor, Docent Taina Jaatinen, for sharing her vast knowledge on stem cell biology, but also her enthusiasm, optimism, great patience and motivation that helped to guide me through this learning process. Very warm thanks to my supervisor, Docent Jukka Partanen, for support, numerous interesting discussions when planning the studies and all the helpful comments throughout the PhD work.

I am indebted to the reviewers, Docents Heli Skottman and Timo Tuuri, for their time, constructive criticism and valuable comments. I warmly thank Professor Riitta Lahesmaa for her kind promise to act as an opponent for the public defense of this thesis.

Acknowledgement and sincere appreciation go to Docent Jarmo Laine for his encouragement, support and valuable advice in clinical aspects, Docent Sampsa Hautaniemi for offering his expertise in data analysis, Jari Niemi for his enthusiasm in statistical analyses, Tero Satomaa and Doctor Jari Natunen for sharing their endless knowledge on glycobiology, Olli Aitio for his skillful nuclear magnetic resonance analysis, Doctor Daniel Nicorici for his flexibility in performing data analysis, Annamari Heiskanen for her timely mass spectrometric analysis and Maria Blomqvist for her work on glycan structure analysis. Additionally, Professor Olli Yli-Harja and Juhani Saarinen are acknowledged for their fruitful collaboration. My special thanks to the undergraduate student Erkka Valo for his work on data reanalysis.

I wish to thank Docent Riitta Kekomäki and the whole staff of the Finnish Red Cross Cord Blood Bank for providing CB units to this study. I am grateful to ‘STR’ and all colleagues and technicians who shared their knowledge to help carry out this study, notably; Sirkka Mannelin, Doctor Tuija Kekarainen, Noora Alakulppi, Anita Laitinen, Tanja Kaartinen, Doctor
Johanna Nystedt and Docent Leena Valmu. Doctor Hannu Turpeinen ‘transplantation specialist’, Katri Haimila, Doctor Pekka Aroviita and Doctor Suvi Natunen ‘our own glycobiologist’ are thanked for their time and productive input.

My colleagues deserve heartfelt thanks for creating a most enjoyable and variable working atmosphere. I also thank my boss, Research and Development Manager Saara Laitinen, for her sincere care and understanding.

I address special thanks to Maarit Koutonen of Lähituki for the fast and effective support, and to both Marja-Leena Hyvönen and Maija Ekholm for excellent library services. Kevin Hanley is warmly thanked for flexibly organizing the language revision of this thesis and I am grateful for the always warm and efficient assistance of Pirjo Nick.

My very sincere thanks to all my friends for their invaluable companionship and the enriching life experiences we enjoy, providing balance to life inside the laboratory. Thank you all for being there for me.

I owe my deepest gratitude to my mother and father for their love and support throughout my life and for my dear little sister Nina who has been looking after the big sister, even the times when it should be the other way around. I am grateful to my grandparents for being part of my life. My sweetest thanks are to my husband Marko for his love, presence and endless support.

January 2008, Helsinki, Finland
REFERENCES


2. de Bruijn MF, Speck NA, Peeters MC, Dzierzak E. Definitive hematopoietic stem cells first develop within the major arterial regions of the mouse embryo. EMBO J 2000;19:2465-2474.


73. Encabo A, Solves P, Carbonell-Uberos F, Minana MD. The functional immaturity of dendritic cells can be relevant to increased tolerance associated with cord blood transplantation. Transfusion 2007;47:272-279.


REFERENCES


REFERENCES


125. Tulp A, Barnhoorn M, Bause E, Ploegh H. Inhibition of N-linked oligosaccharide trimming mannosidases blocks human B cell development. EMBO J 1986;5:1783-1790.


143. Domino SE, Zhang L, Gillespie PJ, Saunders TL, Lowe JB. Deficiency of reproductive tract alpha(1,2)fucosylated glycans and normal fertility in mice with targeted deletions of the FUT1 or FUT2 alpha(1,2)fucosyltransferase locus. Mol Cell Biol 2001;21:8336-8345.


