PhD
Program in Translational
and Molecular Medicine
DIMET

(XXII cycle, academic year 2008-2009)

University of Milano-Bicocca
School of Medicine and Faculty of Science

Identification of new hematopoietic stem cell subsets with a polyclonal antibody library specific for poorly characterized proteins

Coordinator: Prof. Andrea Biondi

Tutor: Dr. Monica Moro

Dr. Pasquale Creo **Matr. No. 708362**

The research presented in this thesis was performed at INGM-National Institute of Molecular Genetics, Milan, headed by Prof. Sergio Abrignani, in collaboration with Transfusional Department - IRCCS Ospedale Maggiore Policlinico, Mangiagalli e Regina Elena, Milan, and the Oncohematology Unit of the Hospital San Matteo, in Pavia

Table of contents

Chapter 1

General Introduction

The importance of cell subsets

The identification of phenotypically distinct cell fractions within apparently homogeneous cell populations is a key step toward the identification and functional characterization of new cell subsets that often have both peculiar effector functions and specific differentiation pathways. Immunology is one of the best examples of this postulation. Since the discovery of the main lymphocyte strand in the late 70s, to the last discoveries among the lowest regulatory subsets, such as Treg or Th17, every time a new immune system subset was isolated, a significant improvement in the understanding of immunological mechanisms was obtained. From the discovery of the main T lymphocyte subsets in the 1970s[1] to the recent identification of the poorly represented regulatory subsets such as Treg and Th17[2-5], every time that a new T cell subset has been characterized phenotypically, a significant improvement in the understandings of the effector functions of the immune system has been subsequently achieved.

New surface markers identification: classical and new approaches

All the human genome has been sequenced[6] and annotated, and a

significant amount of gene products have been studied in some details. However, the distribution and function of a sizable fraction of human gene products is still poorly known[7]. Generally, in the present post-genomic time, the identification of new proteins on cells of interest has resulted either from classical proteomics approaches[8-10] or from gene expression profile analyses[11,12].

Proteomics is a large-scale study of proteins, in particular their structure and function[13,14]. The term "proteomics" was first coined in 1997[15], to make an analogy with "genomics", the study of the genes. Its field of action is the "proteome", that is the entire complement or proteins, including the modification made to a particular set of proteins[16]. Classical proteomics approaches involve the use of techniques that, starting from a tissue homogenate or a cell lysate, are capable to purify or concentrate proteins, to detect specific proteins (Western blot, ELISA, 2D gel electrophoresis, protein microarrays, mass spectrometry), to determine protein-protein interaction (2 hybrid system, immunoaffinity chromatography), to analyze composition and structure of unknown proteins (mass spectrometry, dual polarization interferometry).

All these techniques could be used to identify new biomarkers.

Western blot (or "protein immunoblot") uses gel electrophoresis to separate native or denatured proteins by molecular weight (denatured conditions), or by the 3D structure of the protein (native/non denaturing conditions, SDS-PAGE). Proteins are then

transferred to a membrane (typically nitrocellulose or PVDF), where they are detected using antibodies specific to the target proteins[17,18]. Antibodies can be detected with chemoluminescent, radioactive or fluorescent methods, or by secondary probing.

An evolution of this system is the 2D-gel electrophoresis. With this approach proteins migrate not only on the basis of the molecular weight but also of the isoelectric point. In this case, protein detection is usually obtained by silver or coomassie staining (and UV revelation, in the first case). Here the protein discrimination rely on the relative unlikeliness that two molecules are similar in two distinct proprieties.

Enzyme-Linked Immunosorbent Assay (ELISA), and comparable assays), are biochemical techniques used mainly in immunology to detect the presence of an antibody or an antigen in a sample. Their share the base principle with the Western blot's one. In ELISAs protein of interest is eventually recognized and bonded at the bottom of a microtiter plate by antibodies (if we are screening antigens) or antigens (for antibodies screening). Excess of protein is washed away, and presence of the bonded proteins is revealed staining with specific antibodies conjugated with enzymatic (or fluorescent) reporter[19-21].

Mass spectrometry (MS) is an analytical technique for the determination of the elemental composition of a sample. It is also used for elucidating the chemical structures of molecules, such as peptides and other chemical compounds. The MS principle

consists of ionizing chemical compounds to generate charged molecules or molecule fragments and measurement of their mass-to-charge ratios, a characteristic typical of a given compound. Checking these ratios on a database is possible to recognize proteins, but only using particular precautions. Biomolecules resulted quite fragile when ionized, and a soft ionization technique like the Matrix-assisted laser desorption/ionization (MALDI), based on the use of a special matrix to protect these molecules is needed. Moreover, the sample usually needs to be concentrated through combination with a chromatographic technique. Complementary approaches to the classic proteomics are the gene expression profile analyses. Analysis of transcriptome, the totality of cellular RNA, and, in particular, of mRNA could provide a large amount of quantitative data about what is actively transcripted in a certain cell or tissue, and could describe a global picture of cellular function. The principal approaches in this field are Real Time Quantitative Polymerase Chain Reaction (RTQ-PCR) and RNA microarrays.

RTQ-PCR is a technique based on the PCR, a biochemical process used to amplify targeted DNA molecules. Its feature is that the amplified DNA is detected as the reaction progresses, in real time, while in standard PCR the reaction product is detected at its end, providing the possibility to quantify the exact concentration of target sequence in the total genome. Combining RTQ-PCR with

reverse transcription, the enzyme-based reaction to transcript an RNA sequence in a double strand DNA homologous, it is possible to quantify a target mRNA (that will be transcribed in a so-called cDNA). These data consent to speculate about the production of a certain protein, as a function of the presence and abundance of its transcript. Detection could be obtained with several methods, and more commons are: non-specific fluorescent dyes that intercalate with any double stranded DNA releasing photon in the process (fluorescence increase in function of the increased DNA double strands); sequence–specific DNA probes consisting of oligonucleotides labeled with a fluorescent reporter and a quencher (probes will bind to a sequence inside our target gene; Polymerase, during the reaction will degrade the probe, freeing the reporter from the quencher)[27-29]

DNA microarray is a multiplex technology more and more used in molecular biology. As proteins array, it is based on a glass chip spotted with different probes, spatially well localized to form a microscopic array. But, in this case, probes are not proteins, but DNA oligonucleotides. Using total cDNA as the target is possible to obtain information on presence and quantity of every transcript for which we set a probe, analyzing the whole transcriptome at one time. The detection is obtained labelling the target DNA sequences with non-specific fluorophores (or silver, or chemoluminescent compounds), and is a direct function of number of bounded target copies[30-34].

All these approaches are sensitive enough to identify new genes and proteins expressed in a given cell population[35-38]. However, they have a series of limitation. First of all, these systems' starting materials are not integral cells, but cell lysates. So it is impossible to assess whether differences in the expression levels of genes or proteins occurs in all of the cells analyzed or in a subset of them. It is therefore difficult to study those cell subsets or lineages that are poorly represented within a population and the amount of starting material, that have to be higher than in our reverse proteomic system, may deeply affects the results obtained with these methods[39]. Moreover, many of these classical techniques show high degree of complexity (MS), or require too much time to be used in a high throughput screening.

We have designed an approach that overcame these limitations. As above-mentioned, the distribution and function of a sizable fraction of human gene products are still poorly known[7]. Part of these proteins is predicted to be transmembrane or secreted, meaning that they may be used by the cell to communicate with the external environment. Starting from this assumption we can expect to find new subset-defining proteins among these poorly known gene products. Undoubtedly, one of the best ways to identify and characterize new proteins is to use specific antibodies. We therefore developed a project aimed at obtaining a polyclonal antibody library composed of individual antisera specific for most of those thousands of poorly known human proteins located outside

the cell.

We selected about 3000 genes potentially encoding for transmembrane proteins so far uncharacterized for distribution and function. These genes were cloned and expressed in *E. coli*. The recombinant proteins so obtained were purified and used to immunize groups of five mice each one. In this way we generated a library of 1639 polyclonal antisera that, in principle, can be used to identify new cell subsets in a chosen cell population.

Our antisera were assessed by flow cytometry on immature or mature hematopoietic cells from healthy donors. This analyses were performed on cord-blood derived Hematopoietic stem cells (HSCs) or on Peripheral Blood Lymphocytes (PBLs) and resulted in the identification of eight new proteins expressed by PBLs subset and of three new proteins expressed on subsets of cord-blood derived HSCs. In this study I was mainly interested in the identification of new markers expressed on HSC demonstrating that our approach is suitable for the study of very poorly represented cell populations, such as HSC subsets within the whole cord blood cell population. Moreover, the use of flow cytometry allows not only to estimate the percentage of cells expressing a given cell surface protein but also to separate live positive cells for further studying phenotypical and functional features of the newly identified population.

Stem cells based therapies: where we are and where we would like to go

Stem cells-based therapies represent a new emerging therapeutic approach to treat a variety of degenerative, neoplastic and genetic diseases.

The term "stem cell" appear in literature since 1896, when Pappenheim used it to describe a precursor cell capable of giving rise to both red and white blood cells[40]. Despite this, the Russian histologist Alexander Maksimov (1874–1928) is usually recognized[41,42] as the creator of the term "stem cell" in 1909[43], at Congress of Hematologic Society in Berlin, postulating the existence a cell able to generate all the haematological lineages. At the beginning the scientific community did not accept this concept that remained neglected for more than 50 years. It was only in the 60s that the existence of stem cells became evident, with the studies of McCulloch and Till[44-46] about the presence of self-renewing cells in mouse Bone Marrow (BM), and with the discovery of adult neurogenesis by Altman[47]. These studies headed to the formulation of a first "stem cell theory" in 1968, by André Gernez[48]. This theory was often revisited during time, and actually the current meaning and use of the term "stem cell" is still under discussion. Two relevant characteristics distinguish stem cells from other cell types. First, they are undifferentiated cells with "self-renewing" abilities. It meant that they are able to keep their

number constant through a very carefully regulated cell division programme. Second, they are able to differentiate in response to both physiological and experimental stimuli, becoming tissue-specific cells able to perform distinctive activities[49]. This proliferation vs. differentiation programme is strictly regulated[50-53]. Indeed, in the so-called "tissues with rapid turn over" specific stem cells, regularly divide to replace dead cells or repair damages. On the other hand, there are tissues where stem cells divide and differentiate only when they receive a specific combination of stimuli, otherwise they remain quiescent.

The regulation of stem cells proliferation vs. differentiation is a key point to understand the potential clinical applications of different types of SCs.

There is a hierarchy in the potential for multi-lineage differentiation of SCs. Fertilized eggs are by definition totipotent by virtue of their ability to orchestrate the formation of an entire organism. However they cannot be considered SCs, because they lack the self-renewing capability[54]. Thus, Embryonic Stem Cells (ESCs) derived from early blastocyst are the most potent of SCs. ESCs are indeed capable of unlimited growth in tissue culture and able to give rise to all cell types of the developing soma (but not the extra embryonic structure such as the placenta). They are therefore defined pluripotent[54]. Pluripotency of hESCs was demonstrated by injection of these cells in immunocompromised mice, where they produce teratomas, encapsulated tumors consisting of

disorganized masses of differentiated tissues from all three embryonic germ layers. This is the most stringent pluripotency proof in use today[55]. hESCs represent therefore an important tool to analyze the relationship between gene function and cell and tissue formation, and may provide a source of cells for transplantation therapies, since they are able to differentiate into almost all tissues. However a strong bioethical debate makes impossible the direct therapeutic use of ESCs in the majority of countries[56-58]. Somatic, or adult, stem cells are progressively more restricted in their potential as well in their self-renewing ability (Figure 1). The term "adult" can generate confusion because it does not indicate the age of the stem cells donor. Adult Stem Cells (ASCs) just means that they are taken after the birth of the donor and not from an embryo. Cord blood derived HSC are indeed adult stem cells.

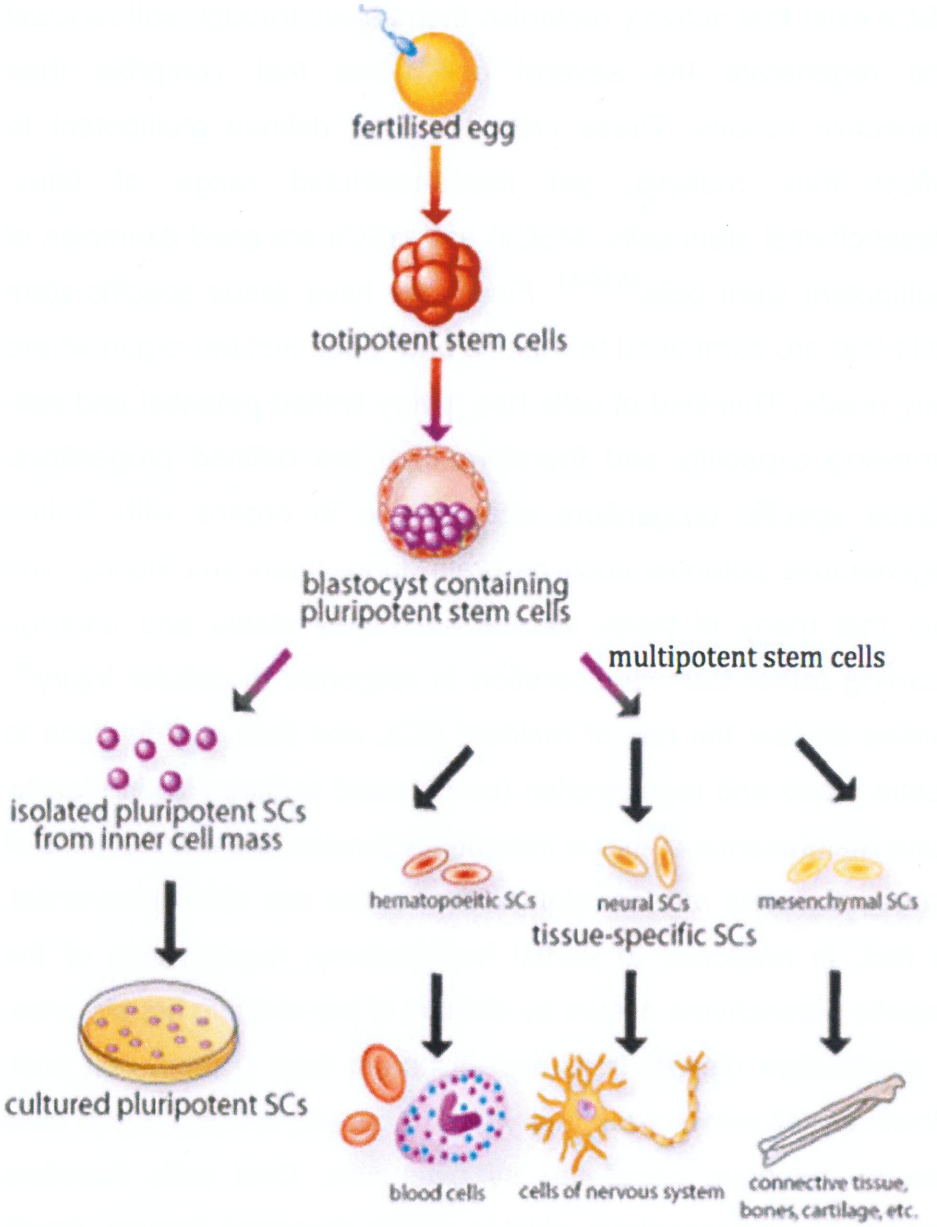

Figure 1. Stem Cells Potency. SCs progressively lost potency during onthogenetic development.

ASCs exist that actively replenish themselves through self-renewal and regenerate the several cell types that comprise their respective tissues. These cells are thus defined multipotent to reflect their multiple, yet tissue-restricted range of fates. Mesenchymal stem cells (MSCs) and HSCs are good examples of multipotent stem cells[53,59-61]. Finally we have tissue specific stem cells that are committed to replace and repair just the organ where they reside. This kind of cells has a very limited potential and self-renewing capability and therefore they are defined progenitors. Tissue specific progenitors often reside in organs with limited regenerative potential such as brain, lung, heart and kidney. The fact that many of these tissues are highly stable and undergo scarring rather than regeneration in response to cellular injury[54], makes unclear the role of resident SCs, and their contributions to tissue repair and regeneration (as opposed to their role in steady-state maintenance of tissue integrity). Likewise, the mechanisms of pancreatic islets regeneration or liver repair are still controversial. In fact, in response to partial hepatectomy, regeneration of the hepatic parenchyma occurs by division of pre-existing hepatocytes, a mechanism distinct from replenishment from a stem cell pool. However, following organ injury, both hepatocytes and ductal cells can be de novo generated by oval cells. Oval cells therefore appear to be progenitors able to repair damaged tissues, tough relatively quiescent in normal conditions[62]. In the same way, muscle stem/satellite cells divide and can regenerate injured

muscle by differentiating into muscle fibers[63]. The regulation mechanisms of somatic stem cells role are under extensive investigation.

The multipotency and, generally speaking, the regenerative potential of stem cells makes them a very promising therapeutic tool. In 1968 was successfully performed the first Bone Marrow Transplantation (BMT) between siblings[64]. From that point scientific and medical communities assisted to a quick escalation of studies, discoveries and clinical trials involving SCs. But a lot of barriers were distributed on that road. First of all, to define the optimal stem cell type to use for regenerating a given tissue is not obvious. A tissue specific progenitor could give the best result, however it is often impossible to get enough cells for transplantation since progenitor cells poorly proliferate in culture. To overcome this limitation Takahashi and Yamanaka created in 2006 the "induced pluripotent SCs" (iPSCs)[65], through ectopic expression of transcriptor factors linked to pluripotency in mouse fibroblasts. Given that iPS cells represent a patient's own genetic make-up, any tissue derived from the line would necessarily be histocompatible, allowing for rejection-proof cell transplantation.

This interesting approach is not without risks. It is always possible that in vitro modified cells go trough neoplastic transformation because of the loose of cell division control. Because of these risks "naturally" pluripotent stem cells represents the best option for therapeutic applications. Multipotent stem cells proliferate a lot in

culture, making easier to get a large amount of cells for transplantation. But in this case to maintain the desired differentiation ability after the in vitro expansion pathway represents a major challenge. To this respect an important consideration is that stem cells subpopulations with different commitments may be phenotypically different. Then, it is likely that they express membrane proteins unknown so far. Thus, the identification of these new stem cell markers would make possible to separate the positive subpopulation, significantly improving their use in therapy. This consideration makes Stem Cells the ideal field for the application of our approach to identify new cell subsets.

HSC subsets: is it everything known?

HSCs are multipotent stem cells present, in different proportion, in the peripheral blood, cord blood and bone marrow[66]. Although the plasticity of HSC is still controversial[67,68], there are evidences indicating that HSC may generate not only blood cells, but also epithelial cells[69,70], endothelial cells[70-72] and hepatocytes[73,74]. HSCs are able to generate daily billions of cells over an entire lifetime[75]. The acknowledged progeny of HSC is extremely heterogeneous, consisting of erythrocytes, platelets derived from megakaryocytes, lymphocytes, granulocytes and macrophages (**Figure 2**). Mature blood cells, with the exception of some rare lymphoid subpopulations, have a relatively short lifespans ranging from few hours (granulocytes) to some weeks (erythrocytes and

lymphocytes). In addition, the balance between different cell types is subjected to rapid changes to cope with different requirements such as bleeding, low oxygen levels or infection[75]. Thus the homeostasis in the blood is a formidable task, and haematopoiesis is likely to be one of the most complex never-ending differentiation processes in adults.

Since the early 50s it was clear that BM transfer was able to rescue radiated mice and guinea pigs[76]. Subsequent experiments suggesting the existence of multipotent progenitor (MPP) cells[45,77] provided support for the idea that cells in the BM are responsible for this radioprotective effect. Further transplantation experiments revealed that the BM contained progenitor cells capable of generating colonies composed of several types of blood cells in the spleen of recipient animals (colony-forming unit spleen, CFU-S)[45,77] suggesting the existence of multipotent hematopoietic progenitor cells. CFU-S cells were also able to give rise to new CFU-S cells upon retransplantation, suggesting that they are able to self-renew[78]. However, later investigations resolved that even though CFU-S forming cells are multipotent, they have a limited ability to sustain blood cell production over time, suggesting that they rather represent MPP, functionally distinct from the hematopoietic stem cells mediating long-term reconstitution of blood cell production in the recipient mice[79]. Further proofs of the existence of a hematopoietic common progenitor were showed in the following years, both molecular[80-82] and cellular, when single

progenitor cells were transplanted in conditioned hosts and assayed for function and lineage potential[83].

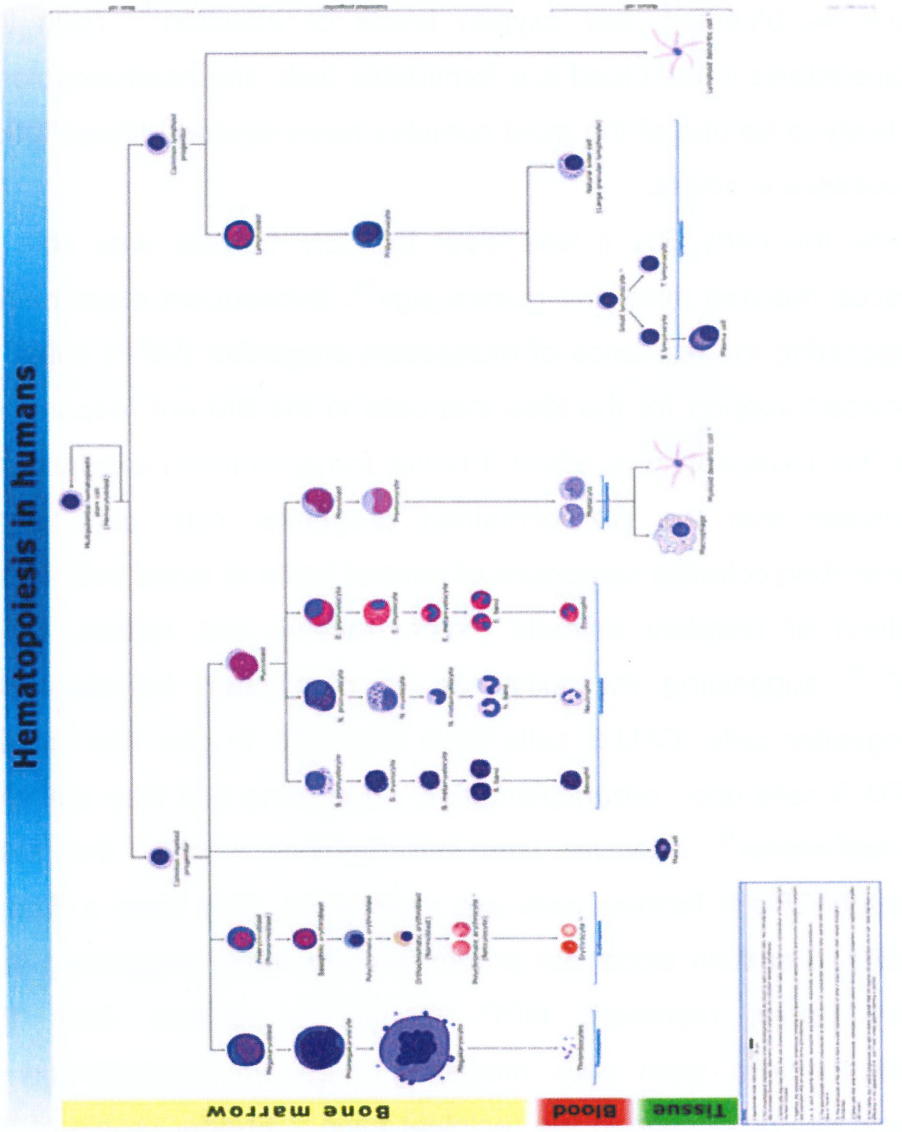

Figure 2. Hematopoiesis. Schematic diagram of hematopoiesis, from HSC to the principal functional lineages.

Phenotypically, HSCs are characterized by the presence of the sialomucin-like adhesion molecule CD34 on their surface[84-86]. The CD34 antigen is expressed on 1–5% of mononuclear BM cells, on a subpopulation of hematopoietic cells, both HSC and early committed progenitors[85]. CD34+ cells have been shown to possess colony-forming potential in short-term assays[86], maintain long-term colony forming potential in *in vitro* cultures[87] and allow the differentiation of blood cell lineages in immunocompromised mice[88]. Initial characterization of HSC was done on the basis of Rh-123 exclusion[89,90] or lectin affinity and showed that CFU-S-8, CFU-S-12 and marrow repopulating cells or pre-CFU-S were physically largely separable[89-91]. In addition, HSC were characterized using their expression of specific surface markers, such as in the mouse with the complete absence of hematopoietic lineage markers (such as CD3, CD14, CD19, CD56, Glycophorine A; a condition we defined as Lin^-, Lineage negative), the expression of the stem cell antigen (Sca-1) and low expression of Thy-1[92-95]. After the successful hematopoietic reconstitution of baboons with selected CD34+ BM, CD34+ cells became the hallmark of murine and human HSC[96]. Donnelly *et al*[97] supported this, demonstrating that murine CD34+ long-term repopulating cells (LRC) are more than 100 times more abundant than CD34− LRC, and that CD34+ cells, not CD34− LRC could be maintained in suspension culture. Although Lin−CD34+ and Lin−CD34− cells contained LRC, he postulated that both constitute two functionally

distinct populations, where in competitive repopulation experiments Lin−CD34+ cells could provide both short- and long-term engraftment, whereas Lin−CD34− cells were only capable of long-term engraftment[97]. Human HSC are currently defined as Lin−CD34+DR− based on results of LTC-IC and various other assays[98-100]. Recent data reveal the presence of highly purified Lin−CD34− subpopulations and suggest the absence of long-term reconstitution potential in the CD34+ fraction. This contrasts with former results on CD34+ HSC, the use of CD34 as a marker for HSC and the long-term effects of CD34+ selection in human transplant settings[101-105].

Nevertheless, since murine studies on CD34− and CD34+ HSC suggest that both are freely interconvertible[106], and if this applies to human HSC also, the CD34+ selection may be appropriate to distinguish potent HSC from quiescent stem cells or mesenchymal precursors. It's clear that HSC have been extensively studied in the last decades, however identification of new HSC subsets would be helpful either to address plasticity questions and to improve their clinical applications. HSC are indeed the only Stem Cells commonly used in therapy. They are also the ideal material for a flow cytometry high throughput screening because of the reasonable accessibility of the starting material (mainly CB) and because they are "naturally" dispersed in solution making unnecessary invasive and complex samples manipulation

Stem cell based therapy: HSC

Allogenic hematopoietic stem cell transplantation (HSCT) is a treatment largely employed for patients affected by a variety of hematological conditions of both malignant and non-malignant origin. Through this procedure, thousands of subjects have been cured from their original disease. Bone marrow (BM) was the first source of HSC successfully used[107,108] and for the last two decades was virtually the sole source of donor cells for HSCT until 1990s. Although this approach kept improving during time, with particular regard to the techniques of HLA-typing, the use of HLA-disparate family donors and the development of adoptive cell therapy strategies, only 50% of the patients in need of HSCT find a suitable HLA-matched donor (related or unrelated) in an acceptable time frame[109,110]. Over the past decade, allogenic cord blood transplantation (CBT) has progressively become a valid alternative for children with both malignant and non-malignant disorders[111-114]. Because of the limited size of the graft product the use of this approach is still limited with adults[115]. Cord blood offers the advantages of easy procurement, the absence of risks to donors, the reduced risk of transmitting infections and, for transplant from unrelated donors, the immediate availability of cryopreserved cells[116,117]. Moreover, mismatches up to two of the six antigens do not preclude the transplant feasibility, as T cells in cord blood are naive and less able, as compare with the counterpart in BM, to cause GVHD[109,118,119]. Possible

disadvantages of CBT are delayed engraftment and higher risk of transplant related mortality (TRM) due to infectious complications[109,120,121]. The higher TRM in CBT patients is related also to the lack of transfer of antigen-experienced T cells, which significantly contribute to the early immunological reconstitution of children given an unmanipulated BMT[118,122]. Haploidentical transplantation of HSC purified from peripheral blood, upon mobilization with growth factors (mobilized peripheral blood stem cells, MPB-SC), represents an immediate alternative to almost all the leukemia patients who fail to find a matched donor or a suitable cord blood unit. The infusion of so-called mega-doses of CD34+ cells, with the concomitant removal of T cells, has been demonstrated to permit a rapid and sustained engraftment, without an increased occurrence of GVHD[110,123,124]. A possible advantage in mismatched transplants is donor-versus-recipient NK cell alloreactivity, which derives from a mismatch between donor NK clones (carrying specific inhibitory receptors for self-HLA class I molecules) and HLA class I ligand on recipient cells[125]. NK alloreactivity may indeed compensate the negative effect of T cell depletion on Graft versus leukemia effect (GVL)[126,127]. The major problem in this kind of transplant is the increased incidence of life-threatening infections, either viral or fungal, as again the recipient cannot benefit of the contribution of adoptively transferred memory T cells[123,128]. The studies performed so far indicated that the immune-recovery and the general outcome are not significantly

different in patients receiving HSCT from these three sources[109,115]. Yet, the majority of these studies focus rather on clinical parameters then on the early dynamics of the immune-system maturation. During my Ph.D. we started a collaboration with the Oncohematology Unit of the Hospital San Matteo, in Pavia, directed by professor Franco Pocatello, to follow the reconstitution of hematopoietic and immune systems in paediatric patients that received an HSCs transplantation in order to cure several malignant and non malignant hematopoiesis-related diseases. HSCs for these patients derive from all the three quoted sources (BM, MPB, CB), and were infused after a total BM deletion. For this reason we hypothesize, due to the young age of the recipients, this system could replicate something very close to the normal hematological onthogenetic development. Characterizing the peculiar cell populations present in the blood of these children, we had the unique opportunity to test our system in a different contest, mirroring our thesis about the role of some of our target proteins in an *ex-vivo* ongoing system. Screening the antisera already resulted positive on HSCs from CB on the specimens from these subjects, we identify at least one interesting circulating subpopulation expressing one of our target protein. This population needs deeper analysis, but could represent advancement in prognosis definition of HSCs Transplantation (HSCT).

Scope of the thesis

The aim of my PhD Project was to develop and certify a new approach for the identifications of surface markers in a specific cell population. In particular, my work was focus on HSCs. Subsets isolated with our "Reverse Proteomics" approach was then characterized for both phenotype and function. On the same time, I would to demonstrate potentiality of our system in the characterization of a totally unknown cell population we isolated in the blood of pediatric patients underwent HSCT.

References

1 Lanzavecchia, A., Maccario, R., Vitiello, A., Nespoli, L. & Ugazio, A. G. Human T lymphocyte subset identification according to avidity for sheep erythrocytes and rosetting capacity with allogeneic red blood cells. *Boll Ist Sieroter Milan* **56**, 217-221 (1977).

2 Dieckmann, D., Plottner, H., Berchtold, S., Berger, T. & Schuler, G. Ex vivo isolation and characterization of CD4(+)CD25(+) T cells with regulatory properties from human blood. *The Journal of experimental medicine* **193**, 1303-1310 (2001).

3 Jonuleit, H. *et al.* Identification and functional characterization of human CD4(+)CD25(+) T cells with regulatory properties isolated from peripheral blood. *The Journal of experimental medicine* **193**, 1285-1294 (2001).

4 Langrish, C. L. *et al.* IL-23 drives a pathogenic T cell population that induces autoimmune inflammation. *The Journal of experimental medicine* **201**, 233-240 (2005).

5 Annunziato, F. *et al.* Phenotypic and functional features of human Th17 cells. *The Journal of experimental medicine* **204**, 1849-1861 (2007).

6 Hocquette, J. F. Where are we in genomics? *J Physiol Pharmacol* **56 Suppl 3**, 37-70 (2005).

7 Haberkorn, U., Altmann, A. & Eisenhut, M. Functional genomics and proteomics--the role of nuclear medicine. *European journal of nuclear medicine and molecular imaging* **29**, 115-132 (2002).

8 Loyet, K. M., Ouyang, W., Eaton, D. L. & Stults, J. T. Proteomic profiling of surface proteins on Th1 and Th2 cells. *Journal of proteome research* **4**, 400-409 (2005).

9 Liu, F. *et al.* Insights into human CD34+ hematopoietic stem/progenitor cells through a systematically proteomic survey coupled with transcriptome. *Proteomics* **6**, 2673-2692 (2006).

10 Foster, L. J. *et al.* Differential expression profiling of membrane proteins by quantitative proteomics in a human mesenchymal stem cell line undergoing osteoblast differentiation. *Stem cells (Dayton, Ohio)* **23**, 1367-1377 (2005).

11 Hemmoranta, H. *et al.* Transcriptional profiling reflects shared

and unique characters for CD34+ and CD133+ cells. *Stem cells and development* **15**, 839-851 (2006).

12 Jaatinen, T. *et al.* Global gene expression profile of human cord blood-derived CD133+ cells. *Stem cells (Dayton, Ohio)* **24**, 631-641 (2006).

13 Anderson, N. L. & Anderson, N. G. Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis* **19**, 1853-1861, doi:10.1002/elps.1150191103 (1998).

14 Twyman, R. M. *Principles Of Proteomics*. (BIOS Scientific Publishers, 2004).

15 James, P. Protein identification in the post-genome era: the rapid rise of proteomics. *Q Rev Biophys* **30**, 279-331 (1997).

16 Wilkins, M. R. *et al.* From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology (N Y)* **14**, 61-65 (1996).

17 Towbin, H., Staehelin, T. & Gordon, J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. 1979. *Biotechnology* **24**, 145-149 (1992).

18 Renart, J., Reiser, J. & Stark, G. R. Transfer of proteins from gels to diazobenzyloxymethyl-paper and detection with antisera: a method for studying antibody specificity and antigen structure. *Proceedings of the National Academy of Sciences of the United States of America* **76**, 3116-3120 (1979).

19 Crowther, J. R. The ELISA guidebook. *Methods in molecular biology (Clifton, N.J* **149**, III-IV, 1-413 (2000).

20 *The Immunoassay Handbook*. 3rd edn, (Elsevier, 2005).

21 Leng, S. X. *et al.* ELISA and multiplex technologies for cytokine measurement in inflammation and aging research. *The journals of gerontology* **63**, 879-884, doi:63/8/879 [pii] (2008).

22 Stoevesandt, O., Taussig, M. J. & He, M. Protein microarrays: high-throughput tools for proteomics. *Expert Rev Proteomics* **6**, 145-157, doi:10.1586/epr.09.2 (2009).

23 Chen, D. S. & Davis, M. M. Molecular and functional analysis using live cell microarrays. *Curr Opin Chem Biol* **10**, 28-34, doi:S1367-5931(06)00002-0 [pii] 10.1016/j.cbpa.2006.01.001 (2006).

24 Jones, R. B., Gordus, A., Krall, J. A. & MacBeath, G. A quantitative protein interaction network for the ErbB receptors using protein microarrays. *Nature* **439**, 168-174, doi:nature04177 [pii]
10.1038/nature04177 (2006).
25 Zhu, H. *et al.* Global analysis of protein activities using proteome chips. *Science (New York, N.Y* **293**, 2101-2105, doi:10.1126/science.1062191
1062191 [pii] (2001).
26 MacBeath, G. & Schreiber, S. L. Printing proteins as microarrays for high-throughput function determination. *Science (New York, N.Y* **289**, 1760-1763, doi:8806 [pii] (2000).
27 Logan, J., Edwards, K. & Saunders, N. *Real-Time PCR: Current Technology and Applications.* . (Caister Academic Press, 2009).
28 Sails, A. D. in *Real-Time PCR: Current Technology and Applications.* (Caister Academic Press, 2009).
29 VanGuilder, H. D., Vrana, K. E. & Freeman, W. M. Twenty-five years of quantitative PCR for gene expression analysis. *BioTechniques* **44**, 619-626, doi:000112776 [pii]
10.2144/000112776 (2008).
30 Pollack, J. R. *et al.* Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat Genet* **23**, 41-46, doi:10.1038/12640 (1999).
31 Ramsay, G. DNA chips: state-of-the art. *Nature biotechnology* **16**, 40-44, doi:10.1038/nbt0198-40 (1998).
32 Schena, M. *et al.* Microarrays: biotechnology's discovery platform for functional genomics. *Trends in biotechnology* **16**, 301-306, doi:S0167-7799(98)01219-0 [pii] (1998).
33 Watson, A., Mazumder, A., Stewart, M. & Balasubramanian, S. Technology for microarray analysis of gene expression. *Curr Opin Biotechnol* **9**, 609-614, doi:S0958-1669(98)80138-9 [pii] (1998).
34 Cheung, V. G. *et al.* Making and reading microarrays. *Nat Genet* **21**, 15-19, doi:10.1038/4439 (1999).
35 Oh, J. M. *et al.* A database of protein expression in lung cancer. *Proteomics* **1**, 1303-1319 (2001).
36 Xu, C. *et al.* Identification and characterization of 177 unreported genes associated with liver regeneration. *Genomics Proteomics*

Bioinformatics **2**, 109-118 (2004).

37 Madras, N., Gibbs, A. L., Zhou, Y., Zandstra, P. W. & Aubin, J. E. Modeling stem cell development by retrospective analysis of gene expression profiles in single progenitor-derived colonies. *Stem cells (Dayton, Ohio)* **20**, 230-240 (2002).

38 Nierhoff, D. *et al.* New cell surface markers for murine fetal hepatic stem cells identified through high density complementary DNA microarrays. *Hepatology (Baltimore, Md* (2007).

39 Lyman, G. H. & Kuderer, N. M. Gene expression profile assays as predictors of recurrence-free survival in early-stage breast cancer: a metaanalysis. *Clin Breast Cancer* **7**, 372-379 (2006).

40 Pappenheim, A. Ueber Entwickelung und Ausbildung der Erythroblasten. *Archiv für Pathologische Anatomie und Physiologie und für Klinische Medicin* **145**, 587-643 (1896).

41 Fliedner, T. M. Prologue to characteristics and potentials of blood stem cells, Stem Cells volume 16, supplement 1. *Stem cells (Dayton, Ohio)* **16**, 357-360, doi:10.1002/stem.160357 (1998).

42 Williams, M. A. L. W. J. *Williams Hematology* 7th edn, (McGraw-Hill, 2006).

43 Maximow, A. Der Lymphozyt als gemeinsame Stammzelle der verschiedenen Blutelemente in der embryonalen Entwicklung und im postfetalen Leben der Säugetiere. *Folia Haematologica* **8**, 125-134 (1909).

44 McCulloch, E. A. & Till, J. E. The radiation sensitivity of normal mouse bone marrow cells, determined by quantitative marrow transplantation into irradiated mice. *Radiation research* **13**, 115-125 (1960).

45 Till, J. E. & Mc, C. E. A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. *Radiation research* **14**, 213-222 (1961).

46 McCulloch, E. A. & Till, J. E. The sensitivity of cells from normal mouse bone marrow to gamma radiation in vitro and in vivo. *Radiation research* **16**, 822-832 (1962).

47 Altman, J. & Das, G. D. Postnatal neurogenesis in the guinea-pig. *Nature* **214**, 1098-1101 (1967).

48 Gernez, A. *Néo-postulats biologiques et pathogéniques.*

(Verschave, 1968).
49 National Institute of Healt, U. S. D. o. H. a. H. S. *Stem Cell Basics: Introduction*, 2009).
50 Firestone, A. J. & Chen, J. K. Controlling destiny through chemistry: small-molecule regulators of cell fate. *ACS Chem Biol* **5**, 15-34, doi:10.1021/cb900249y (2010).
51 Liang, Y., Russell, I., Walworth, C. & Chen, C. Gene expression in stem cells. *Crit Rev Eukaryot Gene Expr* **19**, 289-300, doi:3837c3e737b547ae,3f2ad6af7681b895 [pii] (2009).
52 Jeong, Y. & Mangelsdorf, D. J. Nuclear receptor regulation of stemness and stem cell differentiation. *Exp Mol Med* **41**, 525-537, doi:10.3858/emm.2009.41.8.091 (2009).
53 Kuhn, N. Z. & Tuan, R. S. Regulation of stemness and stem cell niche of mesenchymal stem cells: implications in tumorigenesis and metastasis. *Journal of cellular physiology* **222**, 268-277, doi:10.1002/jcp.21940 (2010).
54 Daley, G. Q. Stem cells: roadmap to the clinic. *The Journal of clinical investigation* **120**, 8-10, doi:41801 [pii] 10.1172/JCI41801 (2010).
55 Lensch, M. W., Schlaeger, T. M., Zon, L. I. & Daley, G. Q. Teratoma formation assays with human embryonic stem cells: a rationale for one type of human-animal chimera. *Cell stem cell* **1**, 253-258, doi:S1934-5909(07)00086-0 [pii] 10.1016/j.stem.2007.07.019 (2007).
56 Lo, B. & Parham, L. Ethical issues in stem cell research. *Endocr Rev* **30**, 204-213, doi:er.2008-0031 [pii] 10.1210/er.2008-0031 (2009).
57 Zarzeczny, A. & Caulfield, T. Emerging ethical, legal and social issues associated with stem cell research & and the current role of the moral status of the embryo. *Stem Cell Rev* **5**, 96-101, doi:10.1007/s12015-009-9062-4 (2009).
58 Vicini, A. Ethical issues and approaches in stem cell research: from international insights to a proposal. *J Soc Christ Ethics* **23**, 71-98 (2003).
59 Bernardo, M. E., Locatelli, F. & Fibbe, W. E. Mesenchymal stromal cells. *Annals of the New York Academy of Sciences* **1176**, 101-117, doi:NYAS4607 [pii]

10.1111/j.1749-6632.2009.04607.x (2009).
60 Oguro, H. *et al.* Poised lineage specification in multipotential hematopoietic stem and progenitor cells by the polycomb protein Bmi1. *Cell stem cell* **6**, 279-286, doi:S1934-5909(10)00006-8 [pii] 10.1016/j.stem.2010.01.005 (2010).
61 Benveniste, P. *et al.* Intermediate-term hematopoietic stem cells with extended but time-limited reconstitution potential. *Cell stem cell* **6**, 48-58, doi:S1934-5909(09)00619-5 [pii] 10.1016/j.stem.2009.11.014 (2010).
62 Zaret, K. S. & Grompe, M. Generation and regeneration of cells of the liver and pancreas. *Science (New York, N.Y* **322**, 1490-1494, doi:322/5907/1490 [pii] 10.1126/science.1161431 (2008).
63 Cerletti, M., Shadrach, J. L., Jurga, S., Sherwood, R. & Wagers, A. J. Regulation and function of skeletal muscle stem cells. *Cold Spring Harbor symposia on quantitative biology* **73**, 317-322, doi:sqb.2008.73.054 [pii] 10.1101/sqb.2008.73.054 (2008).
64 Cooper, M. D. In memoriam. Robert A. Good May 21, 1922-June 13, 2003. *J Immunol* **171**, 6318-6319 (2003).
65 Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676, doi:S0092-8674(06)00976-7 [pii] 10.1016/j.cell.2006.07.024 (2006).
66 Petit, I. *et al.* G-CSF induces stem cell mobilization by decreasing bone marrow SDF-1 and up-regulating CXCR4. *Nature immunology* **3**, 687-694, doi:10.1038/ni813 ni813 [pii] (2002).
67 Masson, S., Harrison, D. J., Plevris, J. N. & Newsome, P. N. Potential of hematopoietic stem cell therapy in hepatology: a critical review. *Stem cells (Dayton, Ohio)* **22**, 897-907 (2004).
68 Wagers, A. J., Sherwood, R. I., Christensen, J. L. & Weissman, I. L. Little evidence for developmental plasticity of adult hematopoietic stem cells. *Science (New York, N.Y* **297**, 2256-2259 (2002).
69 Harris, R. G. *et al.* Lack of a fusion requirement for development of bone marrow-derived epithelia. *Science (New York, N.Y* **305**,

90-93 (2004).
70 Cogle, C. R. *et al.* Adult human hematopoietic cells provide functional hemangioblast activity. *Blood* **103**, 133-135 (2004).
71 Grant, M. B. *et al.* Adult hematopoietic stem cells provide functional hemangioblast activity during retinal neovascularization. *Nature medicine* **8**, 607-612 (2002).
72 Harris, J. R. *et al.* Bone marrow-derived cells home to and regenerate retinal pigment epithelium after injury. *Investigative ophthalmology & visual science* **47**, 2108-2113 (2006).
73 Petersen, B. E. *et al.* Bone marrow as a potential source of hepatic oval cells. *Science (New York, N.Y* **284**, 1168-1170 (1999).
74 Jang, Y. Y., Collector, M. I., Baylin, S. B., Diehl, A. M. & Sharkis, S. J. Hematopoietic stem cells convert into liver cells within days without fusion. *Nature cell biology* **6**, 532-539 (2004).
75 Ogawa, M. Differentiation and proliferation of hematopoietic stem cells. *Blood* **81**, 2844-2853 (1993).
76 Lorenz, E., Uphoff, D., Reid, T. R. & Shelton, E. Modification of irradiation injury in mice and guinea pigs by bone marrow injections. *Journal of the National Cancer Institute* **12**, 197-201 (1951).
77 Becker, A. J., Mc, C. E. & Till, J. E. Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. *Nature* **197**, 452-454 (1963).
78 Juraskova, V. & Tkadlecek, L. Character of primary and secondary colonies of haematopoiesis in the spleen of irradiated mice. *Nature* **206**, 951-952 (1965).
79 Jones, R. J., Wagner, J. E., Celano, P., Zicha, M. S. & Sharkis, S. J. Separation of pluripotent haematopoietic stem cells from spleen colony-forming cells. *Nature* **347**, 188-189, doi:10.1038/347188a0 (1990).
80 Lemischka, I. R., Raulet, D. H. & Mulligan, R. C. Developmental potential and dynamic behavior of hematopoietic stem cells. *Cell* **45**, 917-927, doi:0092-8674(86)90566-0 [pii] (1986).
81 Keller, G., Paige, C., Gilboa, E. & Wagner, E. F. Expression of a foreign gene in myeloid and lymphoid cells derived from multipotent haematopoietic precursors. *Nature* **318**, 149-154 (1985).

82 Dick, J. E., Magli, M. C., Huszar, D., Phillips, R. A. & Bernstein, A. Introduction of a selectable gene into primitive stem cells capable of long-term reconstitution of the hemopoietic system of W/Wv mice. *Cell* **42**, 71-79, doi:S0092-8674(85)80102-1 [pii] (1985).

83 Smith, L. G., Weissman, I. L. & Heimfeld, S. Clonal analysis of hematopoietic stem-cell differentiation in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 2788-2792 (1991).

84 Andrews, R. G., Singer, J. W. & Bernstein, I. D. Precursors of colony-forming cells in humans can be distinguished from colony-forming cells by expression of the CD33 and CD34 antigens and light scatter properties. *The Journal of experimental medicine* **169**, 1721-1731 (1989).

85 Krause, D. S., Fackler, M. J., Civin, C. I. & May, W. S. CD34: structure, biology, and clinical utility. *Blood* **87**, 1-13 (1996).

86 Civin, C. I. *et al.* Antigenic analysis of hematopoiesis. III. A hematopoietic progenitor cell surface antigen defined by a monoclonal antibody raised against KG-1a cells. *J Immunol* **133**, 157-165 (1984).

87 Sutherland, H. J., Eaves, C. J., Eaves, A. C., Dragowska, W. & Lansdorp, P. M. Characterization and partial purification of human marrow cells capable of initiating long-term hematopoiesis in vitro. *Blood* **74**, 1563-1570 (1989).

88 Bhatia, M., Wang, J. C., Kapp, U., Bonnet, D. & Dick, J. E. Purification of primitive human hematopoietic cells capable of repopulating immune-deficient mice. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 5320-5325 (1997).

89 Ploemacher, R. E. & Brons, R. H. Separation of CFU-S from primitive cells responsible for reconstitution of the bone marrow hemopoietic stem cell compartment following irradiation: evidence for a pre-CFU-S cell. *Experimental hematology* **17**, 263-266 (1989).

90 Testa, N. G. & Molineux, G. *Haematopoiesis. A Practical Approach.*, (IRL Press at Oxford University Press, 1993).

91 Ploemacher, R. E. & Brons, N. H. Cells with marrow and spleen

repopulating ability and forming spleen colonies on day 16, 12, and 8 are sequentially ordered on the basis of increasing rhodamine 123 retention. *Journal of cellular physiology* **136**, 531-536, doi:10.1002/jcp.1041360320 (1988).

92 Orlic, D. & Bodine, D. M. What defines a pluripotent hematopoietic stem cell (PHSC): will the real PHSC please stand up! *Blood* **84**, 3991-3994 (1994).

93 Spangrude, G. J., Heimfeld, S. & Weissman, I. L. Purification and characterization of mouse hematopoietic stem cells. *Science (New York, N.Y* **241**, 58-62 (1988).

94 Uchida, N. & Weissman, I. L. Searching for hematopoietic stem cells: evidence that Thy-1.1lo Lin- Sca-1+ cells are the only stem cells in C57BL/Ka-Thy-1.1 bone marrow. *The Journal of experimental medicine* **175**, 175-184 (1992).

95 Krause, D. S. *et al.* Characterization of murine CD34, a marker for hematopoietic progenitor and stem cells. *Blood* **84**, 691-701 (1994).

96 Berenson, R. J. *et al.* Antigen CD34+ marrow cells engraft lethally irradiated baboons. *The Journal of clinical investigation* **81**, 951-955, doi:10.1172/JCI113409 (1988).

97 Donnelly, D. S., Zelterman, D., Sharkis, S. & Krause, D. S. Functional activity of murine CD34+ and CD34- hematopoietic stem cell populations. *Experimental hematology* **27**, 788-796, doi:S0301-472X(99)00032-6 [pii] (1999).

98 DiGiusto, D. *et al.* Human fetal bone marrow early progenitors for T, B, and myeloid cells are found exclusively in the population expressing high levels of CD34. *Blood* **84**, 421-432 (1994).

99 Cashman, J. D. *et al.* Kinetic evidence of the regeneration of multilineage hematopoiesis from primitive cells in normal human bone marrow transplanted into immunodeficient mice. *Blood* **89**, 4307-4316 (1997).

100 Hogan, C. J. *et al.* Engraftment and development of human CD34(+)-enriched cells from umbilical cord blood in NOD/LtSz-scid/scid mice. *Blood* **90**, 85-96 (1997).

101 Goodell, M. A. Introduction: Focus on hematology. CD34(+) or CD34(-): does it really matter? *Blood* **94**, 2545-2547 (1999).

102 Ogawa, M. Changing phenotypes of hematopoietic stem cells.

Experimental hematology **30**, 3-6, doi:S0301472X01007706 [pii] (2002).

103 Dao, M. A. & Nolta, J. A. CD34: to select or not to select? That is the question. *Leukemia* **14**, 773-776 (2000).

104 Andrews, R. G. *et al.* Differential engraftment of genetically modified CD34(+) and CD34(-) hematopoietic cell subsets in lethally irradiated baboons. *Experimental hematology* **28**, 508-518, doi:S0301-472X(00)00137-5 [pii] (2000).

105 Goodell, M. A., Brose, K., Paradis, G., Conner, A. S. & Mulligan, R. C. Isolation and functional properties of murine hematopoietic stem cells that are replicating in vivo. *The Journal of experimental medicine* **183**, 1797-1806 (1996).

106 Sato, T., Laver, J. H. & Ogawa, M. Reversible expression of CD34 by murine hematopoietic stem cells. *Blood* **94**, 2548-2554 (1999).

107 Gatti, R. A., Meuwissen, H. J., Allen, H. D., Hong, R. & Good, R. A. Immunological reconstitution of sex-linked lymphopenic immunological deficiency. *Lancet* **2**, 1366-1369 (1968).

108 Bach, F. H., Albertini, R. J., Joo, P., Anderson, J. L. & Bortin, M. M. Bone-marrow transplantation in a patient with the Wiskott-Aldrich syndrome. *Lancet* **2**, 1364-1366 (1968).

109 Locatelli, F. *et al.* The changing role of stem cell transplantation in childhood. *Bone marrow transplantation* **41 Suppl 2**, S3-7 (2008).

110 Dey, B. R. & Spitzer, T. R. Current status of haploidentical stem cell transplantation. *British journal of haematology* **135**, 423-437 (2006).

111 Wagner, J. E., Kernan, N. A., Steinbuch, M., Broxmeyer, H. E. & Gluckman, E. Allogeneic sibling umbilical-cord-blood transplantation in children with malignant and non-malignant disease. *Lancet* **346**, 214-219 (1995).

112 Kurtzberg, J. *et al.* Placental blood as a source of hematopoietic stem cells for transplantation into unrelated recipients. *The New England journal of medicine* **335**, 157-166 (1996).

113 Gluckman, E. *et al.* Outcome of cord-blood transplantation from related and unrelated donors. Eurocord Transplant Group and the European Blood and Marrow Transplantation Group. *The New England journal of medicine* **337**, 373-381 (1997).

114 Moretta, A. *et al.* Analysis of immune reconstitution in children

undergoing cord blood transplantation. *Experimental hematology* **29**, 371-379 (2001).

115 Welniak, L. A., Blazar, B. R. & Murphy, W. J. Immunobiology of allogeneic hematopoietic stem cell transplantation. *Annual review of immunology* **25**, 139-170 (2007).

116 Gluckman, E. G., Roch, V. V. & Chastang, C. Use of Cord Blood Cells for Banking and Transplant. *Oncologist* **2**, 340-343 (1997).

117 Gluckman, E. & Rocha, V. History of the clinical use of umbilical cord blood hematopoietic cells. *Cytotherapy* **7**, 219-227 (2005).

118 Rocha, V. *et al.* Comparison of outcomes of unrelated bone marrow and umbilical cord blood transplants in children with acute leukemia. *Blood* **97**, 2962-2971 (2001).

119 Cohen, S. B., Perez-Cruz, I., Fallen, P., Gluckman, E. & Madrigal, J. A. Analysis of the cytokine production by cord and adult blood. *Human immunology* **60**, 331-336 (1999).

120 Gluckman, E. *et al.* Factors associated with outcomes of unrelated cord blood transplant: guidelines for donor choice. *Experimental hematology* **32**, 397-407 (2004).

121 Michel, G. *et al.* Unrelated cord blood transplantation for childhood acute myeloid leukemia: a Eurocord Group analysis. *Blood* **102**, 4290-4297 (2003).

122 Eapen, M. *et al.* Outcomes of transplantation of unrelated donor umbilical cord blood and bone marrow in children with acute leukaemia: a comparison study. *Lancet* **369**, 1947-1954 (2007).

123 Aversa, F. *et al.* Treatment of high-risk acute leukemia with T-cell-depleted stem cells from related donors with one fully mismatched HLA haplotype. *The New England journal of medicine* **339**, 1186-1193 (1998).

124 Platzbecker, U., Ehninger, G. & Bornhauser, M. Allogeneic transplantation of CD34+ selected hematopoietic cells--clinical problems and current challenges. *Leukemia & lymphoma* **45**, 447-453 (2004).

125 Velardi, A., Ruggeri, L., Alessandro, Moretta & Moretta, L. NK cells: a lesson from mismatched hematopoietic transplantation. *Trends in immunology* **23**, 438-444 (2002).

126 Ruggeri, L. *et al.* Effectiveness of donor natural killer cell alloreactivity in mismatched hematopoietic transplants. *Science*

(New York, N.Y **295**, 2097-2100 (2002).
127 Farag, S. S., Fehniger, T. A., Ruggeri, L., Velardi, A. & Caligiuri, M. A. Natural killer cell receptors: new biology and insights into the graft-versus-leukemia effect. *Blood* **100**, 1935-1947 (2002).
128 Aversa, F. *et al.* Full haplotype-mismatched hematopoietic stem-cell transplantation: a phase II study in patients with acute leukemia at high risk of relapse. *J Clin Oncol* **23**, 3447-3454 (2005).

Chapter 2

Identification of new hematopoietic stem cell subsets with a polyclonal antibody library specific for poorly characterized proteins

Submitted paper

Identification of new hematopoietic cell subsets with a polyclonal antibody library specific for neglected proteins

Monica Moro*[1], Mariacristina Crosti*[1], Pasquale Creo*, Pierangela Gallina*, Serena Curti*, Elisa Sugliano*, Rossana Scavelli*, Davide Cattaneo§, Elena Canidio§, Paolo Sarmientos§, Giuseppe Viale°, Massimiliano Pagani§ and Sergio Abrignani*.

* *Fondazione Istiuto Nazionale di Genetica Molecolare, INGM, via Francesco Sforza 28, 20122 Milano.*
§ PRIMM, Via Olgettina 58, 10132 Milano.
°European Institute of Oncology, IEO, Milan, Italy;

1 These authors contributed equally to this work.

Corresponding authors: Sergio Abrignani (abrignani@ingm.it) and Massimiliano Pagani (massimiliano.pagani@primm.it)

Acknowledgments

This work was supported by the Italian Ministry of Health and partially by a grant from the CARIPLO Foundation

Abstract

We have addressed the issue of identifying new cell specific markers with a reverse proteomic approach whereby approximately 1700 human open reading frames encoding proteins predicted to be transmembrane or secreted have been selected in silico for being poorly known, cloned and expressed in bacteria. These proteins have been purified and used to immunize mice with the aim of obtaining polyclonal antisera mostly specific for linear epitopes. Such a library, made of about 1600 different polyclonal antisera, has been obtained and screened by flow cytometry on cord blood derived hematopoietic stem cells (HSC) and on peripheral blood derived mature lymphocytes (PBLs). We identified three new proteins expressed by fractions of HSCs and eight new proteins expressed by fractions of PBLs. Remarkably, we identified proteins the presence of which had not been demonstrated previously by transcriptomic analysis. From the functional point of view, looking at new proteins expressed on HSCs, we identified one cell surface protein (MOSC-1) the expression of which on a minority of apparently undifferentiated CD34+ HSCs, marks those HSCs that will go toward monocyte/granulocyte differentiation. In conclusion, we show a new way of looking at the membranome by assessing expression of generally neglected proteins with a library of polyclonal antisera, and in so doing we have identified new potential subsets of hematopoietic cells.

Introduction

The identification of phenotypically distinct cell fractions within apparently homogeneous cell populations is a key step toward the identification and functional characterization of new cell subsets that often have both peculiar effector functions and specific differentiation pathways. Immunology offers one of the best example of this assumption. From the discovery of the main T lymphocyte subsets in the 1970s[1] to the recent identification of the poorly represented regulatory subsets such as Treg and Th17[2-5], every time that a new T cell subset has been characterized phenotypically, a significant improvement in the understandings of the effector functions of the immune system has been subsequentely achieved.

All the human genome has been sequenced[6] and annotated, and a significant amount of gene products have been studied in some details. However, the distribution and function of a sizable fraction of human gene products is still poorly known[7]. Generally, in the present post-genomic time, the identification of new proteins on cells of interest has resulted either from classical proteomics approaches[8-10] or from gene expression profile analyses[11,12]. Both these approaches are sensitive enough to identify new genes and proteins expressed in a given cell population[13-16]. However, it is impossible to assess whether differences in the expression levels of genes or proteins occurs in all of the cells analyzed or in a

subset of them. It is therefore difficult to study those cell subsets or lineages that are poorly represented within a population and the amount of starting material may deeply affects the results obtained with these methods[17].

Undoubtedly, one of the best ways to identify and characterize new proteins is to use specific antibodies. We therefore developed a project aimed at obtaining a polyclonal antibody library composed of individual antisera specific for most of those thousands of poorly known human proteins located outside the cell. We focused our attention on those proteins that are predicted in silico to be transmembrane or secreted[18], which have at least a domain predicted to be "outside" the cell and are therefore likely to be used by cells to interact with the external milieu. We assumed that it would have been possible to identify new subset-defining proteins with specific antibodies specific for these poorly known gene products.

We selected *in silico* about 1700 ORFs potentially encoding for membrane proteins so far poorly characterized in distribution and function. These genes have been cloned and expressed in *Escherichia coli*. The recombinant proteins have been purified and used to immunize groups of five mice. We have generated a library of about 1600 (list in **Appendix A**) polyclonal antisera and assessed them by flow cytometry on immature or mature

hematopoietic cells from healthy donors. This analyses were performed on cord-blood derived HSCs or on Peripheral Blood Lymphocytes (PBLs) and resulted in the identification of eight new proteins expressed by PBLs subset and of three new proteins expressed on subsets of cord-blood derived HSCs.

We show that this high-throughput screening is suitable for the study of very poorly represented cell populations, such as HSC subsets within the whole cord blood cell population. Moreover, the use of flow cytometry allows not only to estimate the percentage of cells expressing a given cell surface protein but also to separate live positive cells for further studying phenotypical and functional features of the newly identified population.

Results

Production and validation of the antisera library.

In the human genome there are about 10.000 genes codifying for proteins that are predicted to be either transmembrane or secreted. For about a third of them there is a little information on distribution and function. We have selected in silico about 1700 such genes, expressed them in *E. coli*, purified the proteins and immunized groups of five mice with each individual protein. We thus obtained a library of 1600 mouse antisera specific for poorly

known human secreted or transmembrane proteins. The generation and production of this antisera library is detailed in supplementary material and summarized in figure 1 and will be described in details in a future manuscript (Pagani M. and Sarmientos P., manuscript in preparation).

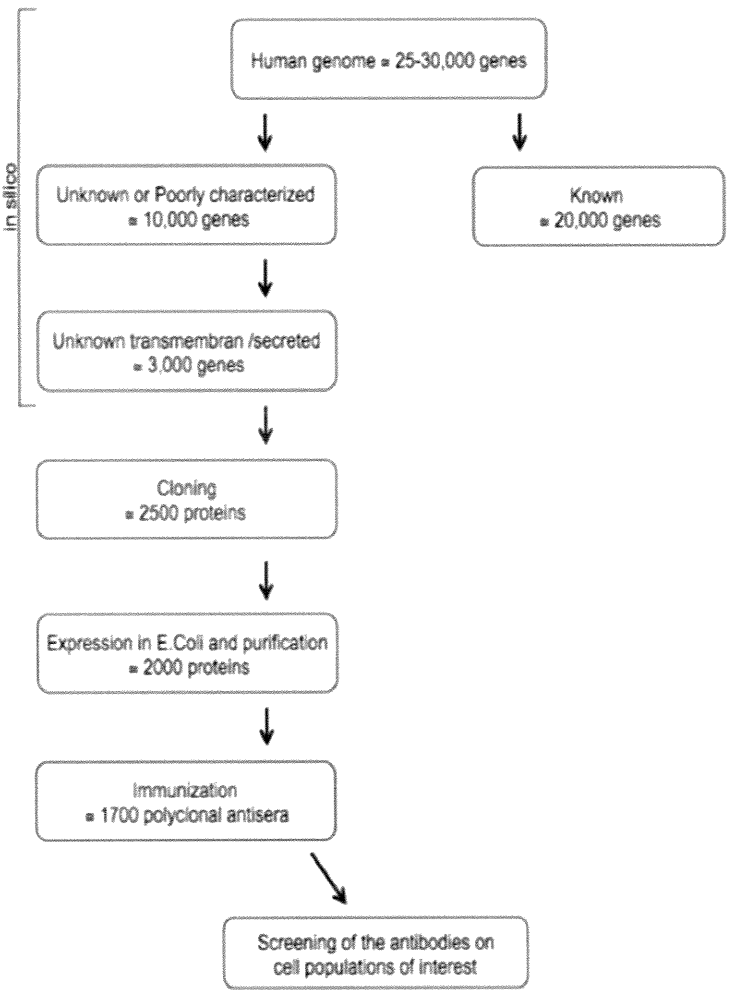

Figure 1. Schematic representation of the antisera library generation.

As these antisera are raised against human proteins expressed in bacteria, most likely they are directed mostly against linear epitopes, a key question was whether such antibodies would recognize the corresponding proteins on the surface of human cells. To obtain proof of concept that our antibody library had the potential to identify new cell surface proteins, we produced in bacteria, with the same technology described above, antisera specific for twenty well characterized proteins (i.e., with assigned CD numbers) known to be present on hematopoietic cells and for which very good monoclonal antibodies exist (**Table 1**). We then asked whether we would be able to "identify" these proteins on living cells by flow cytometry or on fixed and embedded tissues by immunohystochemistry (IHC). **Table 1** summarizes the whole results and shows that with flow cytometry we could "identify" 12/20 (60%) known proteins on the surface of living PBLs. Remarkably, with IHC on fixed and embedded lymph node tissues we "identified" 11/14 (85%) known proteins in lymphoid cells.

SYMBOL	DESCRIPTION	FACS score	IHC
CD2	CD2 (p50), sheep red blood cell receptor	3	positive
CD3δ	CD3 delta polypeptide (TiT3 complex)	0	negative
CD3ε	CD3 epsilon polypeptide (TiT3 complex)	0	positive
CD3γ	CD3 gamma polypeptide (TiT3 complex)	3	negative
CD8α	CD8 alpha polypeptide (p32)	3	positive
CD8β	CD8 beta polypeptide 1 (p37)	0	positive
CD45	Protein tyrosine phosphatase, receptor type, C	0	positive
CD161	Killer cell lectin-like receptor subfamily B, member 1	0	positive
CD25	Interleukin 2 receptor, alpha	3	positive
CD27	Tumor necrosis factor receptor superfamily, member 7	3	positive
CD69	p60, early T-cell activation antigen	1	positive
CD71	Transferrin receptor (p90)	0	positive
CD72	CD72 antigen	3	positive
CD80	CD28 antigen ligand 1, B7-1	3	negative
CD86	CD28 antigen ligand 2, B7-2	3	positive
ITGB7	Integrin, beta 7	0	negative

Table 1. Antisera control list. In the table are indicated all the proteins used as controls. In the columns on the right are showed the results for both FACS and IHC analysis.

Figure 2 shows representative stainings of lymphocytes with these antisera by flow cytometry (**Fig. 2A**) or by IHC (**Fig. 2B**). PBLs were stained with our polyclonal anti-CD8 antiserum and compared it with a commercial anti-CD8 monoclonal antibody (mAb) used alone or in a combination with commercial anti-CD3 and anti-CD4 mAbs (**Fig. 2A**). The percentage of cells identified with the anti-CD8 from our library is comparable to the one obtained with the commercial monoclonal antibody and also the CD4/CD8 ratio among T cells is correctly detected. The representative IHC experiment on lymph nodes in **Fig. 2B** panels **a** and **b** show that the anti CD2 and anti CD3 antisera identified the very great majority of cells in the T cell area with some positive signal in the follicle germinal center, and panels **c** and **d** shows the anti-CD8a and anti-CD8b antisera stained a fraction of cells in the T cell area. Finally, the anti CD72 antiserum, which is B cell specific, identified correctly the follicular area (panel **e**), and the anti-CD69 antiserum (panel **f**) detected, as expected, a relatively small population of recently activated lymphocytes.

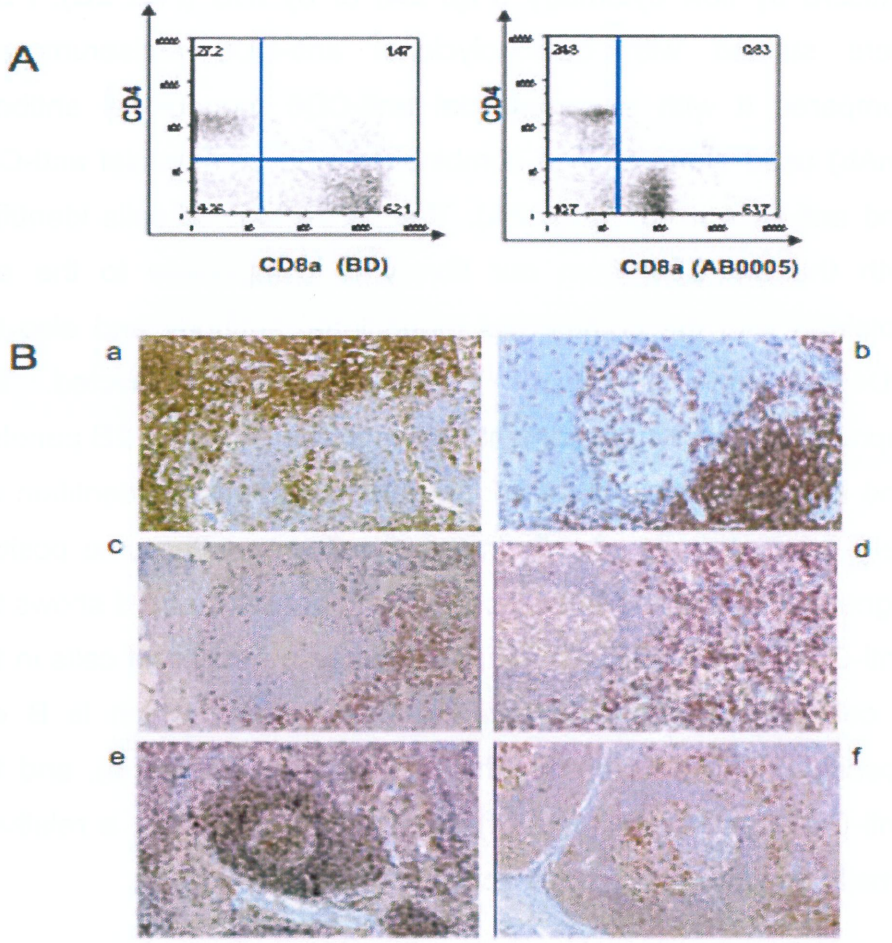

Figure 2. FACS and IHC analysis with sera specific for well-known proteins.
A) Comparison of the CD8 staining performed on PBL with either a commercially available anti CD8 mAb (BD biosciences) or the anti CD8 alpha serum at 1:100 dilution points. Both the samples were stained also with commercially available anti CD3 and anti CD4 mAb (BD biosciences). The distribution of CD4 and CD8 is analyzed upon gating on CD3 positive cells. B) Immunohistochemistry. Sections of Human lymph nodes were pretreated with an antigen retrieval solution and were then incubated with the indicated antisera. Detection steps were done using a commercially available kit according to the manufacturer's instructions. Peroxidase activity was developed with 3-3-diaminobenzidine-copper sulfate to obtain a brown-black end product. a) anti CD2, b) anti CD3 gamma, c) anti CD8 alpha, d) anti CD8 beta, e) anti CD72, f) anti CD69.

From the above results we conclude that our approach is suitable to identify new molecules on a cell population of choice by IHC or flow cytometry, and most importantly that antisera from the library can be used in multi-parametric analysis by flow cytometry. Although the use of IHC would result in a lower number (15%) of false negative antibodies compared to flow cytometry (40%), we decided to utilize Flow cytometry to screen our antibody library on human hematopoietic cells, because of the possibility to gate and analyse fractions of very minor cells subsets (<1-2%) which would pass mostly undetected in a screening performed by IHC.

Identification of new proteins expressed on the surface of PBLs and HSCs

Having validated the approach, we screened our antisera library on resting or activated PBLs and cord blood samples with the aim of identifying molecules that could define new cell substes within PBLs and HSCs. Each individual antiserum was tested in three dilution points on at least three independent PBLs and at least three independent cord blood samples. On PBLs, the screening was performed on 5x10E5 cells from either resting or phytohemagglutinin (PHA)-stimulated lymphocytes. On cord blood samples, 5x10E6 cells were analyzed to eventually gate on 1000-2000 "canonical" HSCs (CD34+, CD45dim). In the search for new cell subsets, we concentrated our efforts on antisera that were

positive for a fraction of the population we were interested in. Antisera that resulted positive with these criteria after the first screening on PBLs or HSCs were validated on cell samples from ten additional independent donors. Finally, to confirm the presence of the transcript corresponding to the protein recognized by the antiserum we assessed mRNAs by RT-PCR analysis.

Figure 3 shows that the high throughput screening of hematopoietic cells with the 1600 antisera led to the identification of seven new molecules expressed on PBL subsets (**Fig. 3A**), of one new molecules up-regulated on a subset of activated PBLs (**Fig. 3B**) and of three new molecules expressed on a subset of HSCs (**Fig. 3C**). To characterize the newly identified cell subsets, we performed a multicolor FACS analysis of the antisera in combination with monoclonal antibodies specific for the known main subsets of PBLs (CD3+ T lymphocytes, CD19+ B lymphocytes, or CD56+ Natural Killer cells) or HSCs (CD34). **Fig. 3A** shows that four out of seven new molecules we identified on PBLs (i.e., LPPR2, MOSC-1, TMEM38B and GSG1L) are mainly present on B lymphocytes, whereas the other three antisera specific for TMCC1, TMEM126B and SUSD3 are positive on both T and B lymphocytes subsets. The presence of the transcripts was confirmed by RT-PCR performed either on total peripheral blood mononuclear cells (**Fig. 3D** panel **a** and **b**) or on HSCs purified

magnetically with an anti-CD34 mAb out of cord blood cells (**Figure 3D** panel **c**).

To rule out the possibility of sera cross-reaction toward unrelated molecules, we cloned, in a mammalian expression vector in frame with a c-myc tag, the genes coding for the putatively identified proteins, transiently transfected HeLa cells, and assessed by western blot protein expression using the specific antiserum or an anti-cmyc mAb. Figure 4 shows that most antisera recognized a band of the expected molecular weight, whereas only two antisera (putatively specific for LPPR2 and GSG1L) did not recognize a band of the expected molecular weight indicating these two antisera could be cross-reactive, originating false positive signals in the screening despite the positive RT-PCR.

From all the above experiments, we conclude that our reverse proteomic approach to address the membranome is specific and sensitive enough to allow the identification of new proteins on hematopoietic cell subsets. Remarkably, we identified proteins the presence of which had not been demonstrated previously by transcriptomic analysis. For Instance, we found (Fig 3C) that a fraction of indifferentiated CD34+ HSCs express MOSC1 on the surface, whereas previous gene expression profile analyses indicated a possible expression of MOSC-1 on peripheral blood monocytes, myeloid hematopoietic precursors but not on HSCs (REF xx http://symatlas.gnf.org/SymAtlas).

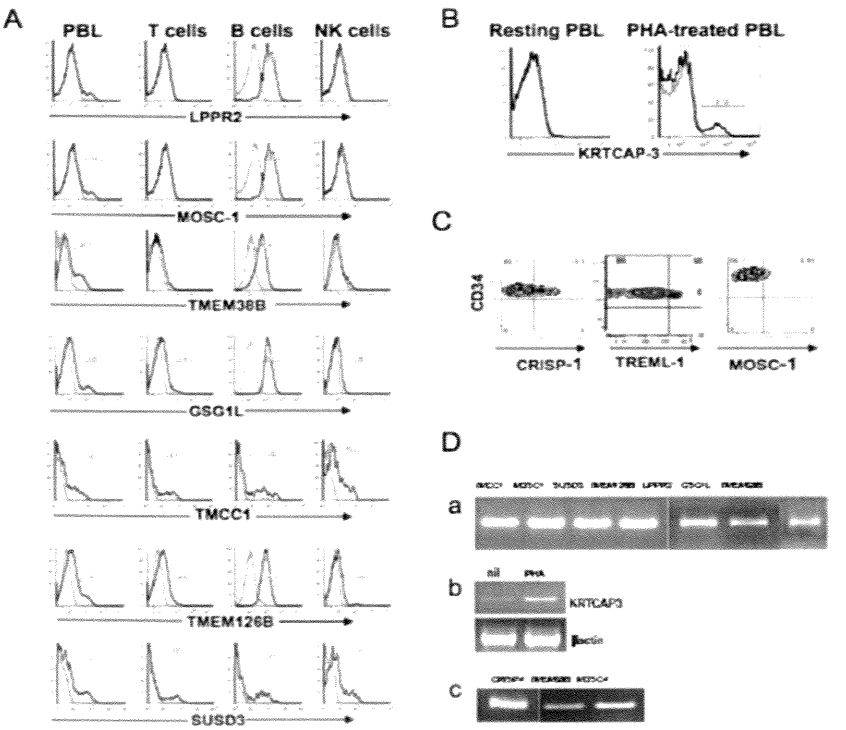

Figure 3. Results of sera screening by FACS on PBL and HSC.
A) FACS analysis of sera positive on PBL after the second level of the screening. PBL were stained with the indicated sera at the optimal dilution point (1:50 to 1:200). The sample were stained also with anti CD3, anti CD19 and anti CD56 mAbs analyze the sera reactivity upon gating on the different subpopulations. A plot representative of five different donors is shown for each serum. B) The KRTCAP-3 specific serum recognizes PHA-treated cells. PBMCs are treated for 24 hours with 1 μg/ml of PHA. After the treatment both unstimulated and treated cells are stained with the KRTCAP-3-specific serum. C) FACS analysis of sera positive of cord blood HSC. Cord blood mononuclear cells are stained with the indicated sera at the optimal concentration (1:50 to 1:100). The samples are stained also with anti CD45 and anti CD34 mAbs to perform the analysis upon gating on CD34highCD45dim (HSC). A plot representative of a least 3 independent donors is shown. D) RT-PCR analysis. A- cDNA from total PBMC were amplified with primers specific for the indicated proteins. b- cDNA from either unstimulated and PHA-treated PBMC was amplified with KRTCAP-3 specific primers. KRTCAP3 expression is up regulated two to three times. Beta actin amplification is used as normalization. c- cDNA samples from HSC were generated by retro-transcription of RNA extracted from a pool of magnetically purified CD34 positive cells from 2-3 independent cord blood units. The purity of the HSC was usually >99%. The samples were amplified with primers specific for the indicated proteins.

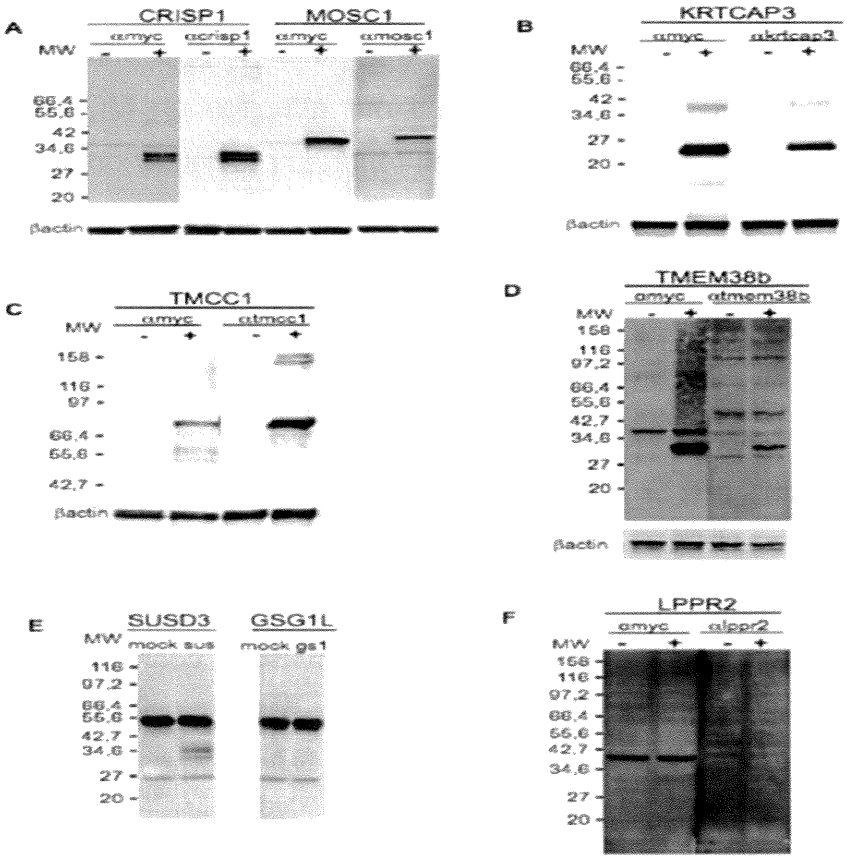

Figure 4. Assessment of antisera specificity on Hela transfected cells.
Hela cells were transiently transfected with a myc-tag version of the proteins identified with the sera library. At 24 hours from the transfection cells were lysated as described in the Method section. 40mg of total proteins were loaded on SDS page and a WB analysis was performer using both an anti myc mAb (9E10 clone) and the corresponding antiserum.
A) WB analysis of Hela cells transfected with CRISP-1 and MOSC-1. In both the cases the anti myc mAb and the specific antiserum recognized a protein of the expected molecular weight that is not present in the cells transfected with the mock vector. A comparable result was obtained wit KRTCAP-3 (B), TMCC-1 (C), TMEM38B (D) and SUSD3 (E) transfected cells.
The WB analysis of GSG1-L cells (E) and LPPR2 cells (F) shows that neither the anti myc nor the specific antiserum is able to recognize in a specific way a protein in transfected cells.

Identification of MOSC1 as a marker of mono-granulocyte development on HSCs.

To address functional aspects associated to the expression of these new proteins, we focused our attention on MOSC-1, the expression of which had not been reported previously on HSCs.

MOSC-1 (Moco Sulphurase domain containing protein-1) is a potentially secreted protein that contains a MOSC domain. This domain is predicted to be a sulfur-carrier domain that receives sulfur abstracted by the pyridoxal phosphate-dependent NifS-like enzymes, on its conserved cysteine, and delivers it for the formation of diverse sulfur-metal clusters[19].

To assess whether the presence of MOSC-1 conferred peculiar functions to the HSCs, we separated MOSC-1 positive HSCs (**Figure 5A** shows a representative plot) by Fluorescence Activated Cell Sorting (FACS) and performed a colony forming cell (CFC) assay to establish the differentiation capacity of CD34+ HSC that expressed or not MOSC-1. Therefore, HSC were plated in a semi-solid medium in the presence of a cocktail of growth factors (SCF, Flt3L, IL-6, GM-CSF, IL-3 and EPO) capable of sustaining proliferation and differentiation of different hematological lineages. After 14 days of culture, erythroyd precursors would generate BFU (red) colonies, myelo-granulocytes precursors would generate CFU-GM, CFU-G, CFU-M (white) colonies and the more immature

precursors would generate CFU-GEMM (mixed) colonies.

Figure 5B shows that MOSC-1 positive HSCs generated almost exclusively white colonies, clearly indicating a commitment of MOSC-1 positive cells toward leukocyte lineages. Moreover, a further phenotypic analysis performed on MOSC-1 positive CD34+ cells (**Figure 5C**) showed CD133 and CD33 co-expression together with the low expression of CD38, CD7 and CD10, thus suggesting possible mono-granulocytic commitment. This possibility is confirmed by the reactivity of MOSC-1 antiserum on peripheral blood monocytes (**Figure 5D**). Although **Figure 5D** shows a MOSC-1-specific reactivity also toward B-lymphocytes, RT-PCR analysis performed on purified PBMC subpopulations (**Figure 5E**) indicated that MOSC-1 transcript is present only in monocytes. Since MOSC-1 is potentially secreted, it is possible that the proteins produced and secreted by monocytes binds a receptor on B cells. In conclusion, the FACS, CFC and RT-PCR results are consistent with an association of MOSC-1 expression on HSCs with a monocyte commitment.

We conclude that using this antibody library we could not only identify new molecules expressed on subsets of cell populations of interest, but could also demonstrate a correlation between a phenotype and the functional commitment of the newly identified cell subset.

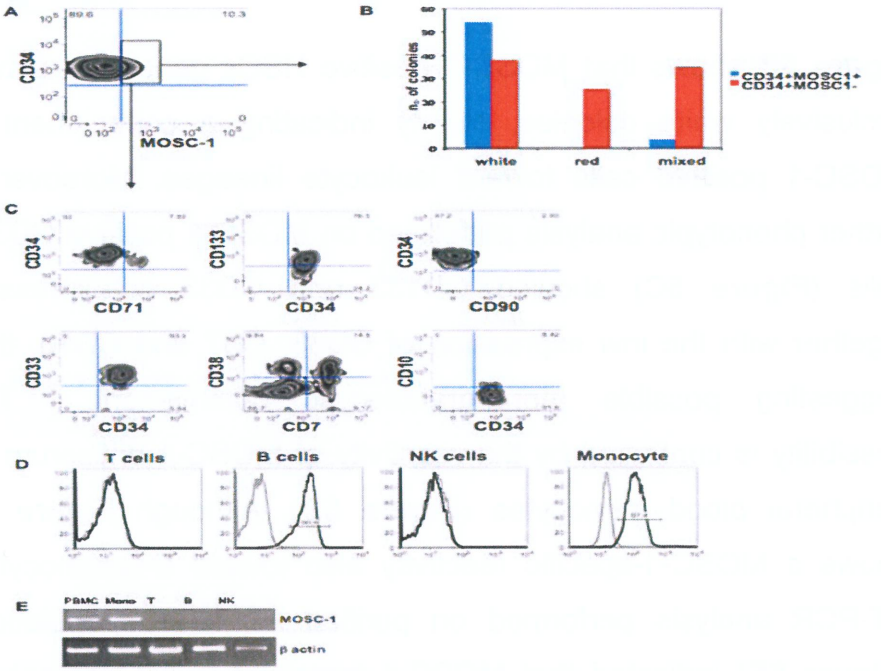

Figure 5. Pre-characterization MOSC-1 expressing HSC.
A) Representative distribution of MOSC-1 on HSC. Cord blood mononuclear cells are stained with MOSC-1-specific serum diluted 1:150. It is shown the analysis upon gating on CD34highCD45dim cells (HSC). B) CFC assay with MOSC-1 positive cells. MOSC-1 positive and MOSC-1 negative HSC were purified by Fluorescence Activated Cell Sorting. The purity of the populations used in the assays was >90%. The same number of cells from the two populations (100-500) were plated in Methocult medium (Stem Cell Tech.) and incubated at 37°C for 14 days. Then the number of white, red and mixed colonies was counted. The average of 2 independent experiments is shown. C) Phenotype of MOSC-1 positive HSC. Cord Blood mononuclear cells were stained with anti MOSC-1 antiserum at 1:150 dilution point after magnetic enrichment of HSC. All the samples were stained also with anti CD34 and anti CD45 mAbs and, in turn, with anti CD71, anti CD33, anti CD133, anti CD90, anti CD38, anti CD7 and anti CD10 mAbs. The expression of these markers is showed on the population of MOSC-1 positive HSC. An analysis representative of three independent experiments is shown. D) MOSC-1 distribution on PBMC. PBMC from healthy donors were stained with anti MOSC-1 antiserum at 1:150 dilution points. The samples were stained also with anti CD3 (T cells), anti CD19 (B ells), anti CD56 (NK cells) and anti CD14 (Monocytes) mAbs to gate the correct subpopulation. An analysis representative of 5 different experiments is shown. E) MOSC-1 RT-PCR on PBMC. RNA from PBMC and from the indicated magnetically purified subpopulation (Purity >99%) was retro-transcribed and amplified with MOSC-1 specific primers. Beta actin gene was amplified as positive control.

Identification of TREML-1 as a marker of megakariocyte development on HSC.

Another very interesting protein we found expressed on a HSC subset is TREML-1 (Triggering Receptor Expressed on Myeloid cells-Like 1). TREML-1 gene is placed in a cluster on chromosome 6 with the single Ig variable (IgV) domain activating receptors TREM1and TREM2 but it has distinct structural and functional properties[20,21]. This protein is a cell surface receptor that enhances calcium signalling in an SHP2-dependent manner, and play a role in the innate and adaptive immune response[22]. TREML-1 was detected in platelets, monocytic leukemia and T-cell leukemia[21-24]. In particular, it resulted sequestered in cytoplasmic vesicles in resting platelets, and transported to the cell surface after thrombin stimulation. Soluble fragments can be released into the serum by proteolysis[23,24].

During our screening, we detected TREML-1 on a variable percentage (4-10%) of HSC (**Figure 6A** is a representative plot). The expression of this protein was confirmed by RT-PCR on both HSC and PBMC (**Figure 6D**).

To investigate a potential commitment of TREML-1 positive HSC toward a specific lineage, TREML-1 expressing cells were isolated by FACS sorting and used in classical CFC assays. As shown in **Figure 6B**, TREML-1 positive cells generate all the possible colonies (BFU, CFU-GM, CFU-G, CFU-M and CFU-GEMM)

indicating that the expression of this protein characterize a quite immature HSC subset.

Since TREML-1 is expressed by mature platelets, we investigated whether its presence on HSC is associated with typical megakariocyte markers, such as CD41 and CD61. We found that half of the TREML-1 positive cells express also CD41 and CD61 (**Figure 6C**) and, on these cells, they seem to be present with a 1:1 ratio (**Figure 6E**). This observation, together with the simultaneous expression of CD133, indicated that TREML-1 could identify an immature HSC subpopulation with a potential commitment towards platelet differentiation.

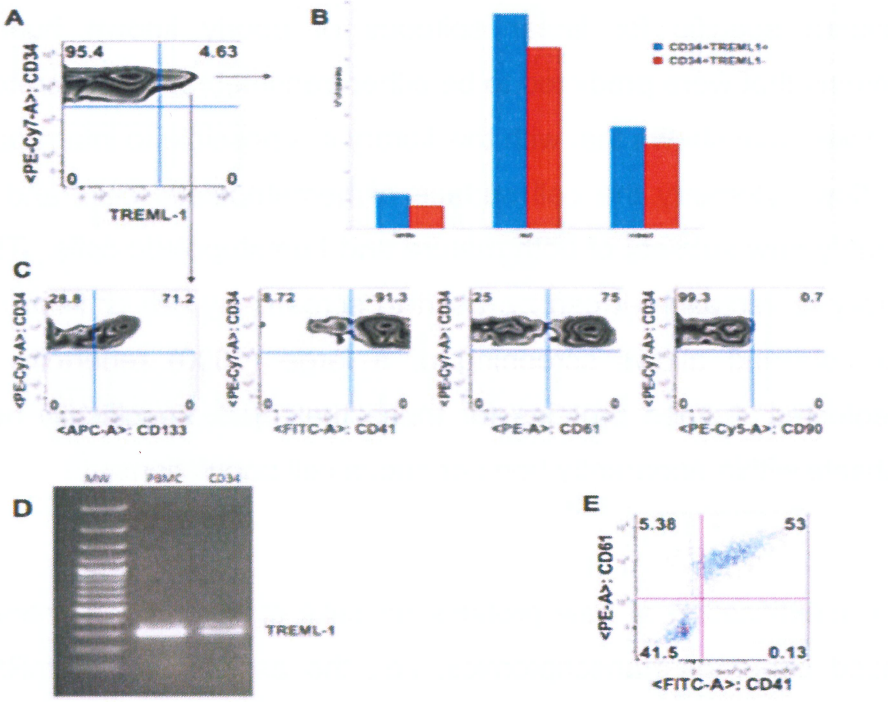

Figure 6. Pre-characterization of TREML-1 expressing HSC.

A) Representative distribution of TREML-1 on HSC. Cord blood mononuclear cells are stained with anti TRML-1 specific polyclonal antiserum diluted 1:150. The analysis was performed gating on D34highCD45dim cells (HSC), B) CFC assay with TREML-1 positive HSC. TREML-1 positive and negative ells were purified by Florescence Activated Cell Sorting, obtaining a purity >90% for the positive population. he same number of cells (100-500) was plated, for both populations, in Methocult semi-solid medium (Stem Cell Tech.) and incubated 37°C for 14 days. The number of white, red and mixed colonies was counted. The average of 3 indipendent experiments is shown. C) Phenotype of TREML-1 positive cells. Cord blood cells were magnetically enriched for HSC and stained with anti TREML-1 specific policlonal antisera diluted 1:150. Cells were stained also with anti CD34 and anti CD45 mAbs and, according to the sample, with anti CD133, anti CD41, anti CD61 and anti CD90. The expression of these markers is showed on the population of MOSC-1 positive HSC. An analysis representative of two independent experiments is shown D) TREML-1 transcript presence demonstrated by RT-PCR in CD34 positive cells and in PBMC. E) CD41 and CD 61 are equally expressed on TREML-1 positive HSC. Plot obtained by gating on TREML-1 positive, CD34highCD45dim HSC from a cord blood unit.

Discussion

In this study we have described a library of mouse polyclonal antisera specific for linear epitopes of poorly known human proteins that were predicted to be either transmebrane or secreted, and we have shown that with this library it is possible to interrogate by flow cytometry the cell surface of hematopoietic cells and to identify new subsets of both mature and hematopoietic cells. This library is versatile—it can be used to screen any cell or tissue of interest—and allows screening of a large (1600) repertoire of "neglected" human proteins for those that mark specifically new subsets within apparently homogeneous cell populations.

The identification of new proteins on cells or tissues is generally based either on transcriptomics, i.e., the assessment of mRNA expression profile, or on proteomics, i.e., the direct identification by mass spectrometry of proteins separated by 2D gels or liquid-based separation methods. Transcriptomics allows to analyze and compare large amount of samples at the same time[7], but poses the problem of the correlation between mRNA levels and protein expression levels. Proteomics is very informative but poses the problem of the complexity of the approach that makes it not suitable for high throughput screenings. In both cases, it is impossible to assess whether differences in the expression levels of genes or proteins occurs in all of the cells analyzed or in a

subset of them. It is therefore difficult to study those cell subsets or lineages that are poorly represented within a population. Our goal was to study hematopoietic cell subsets by flow cytometry and therefore opted for an approach based on the direct identification of proteins with an antibody library.

We elicited our antibody library using proteins expressed in bacteria as immunogens. On the one hand, antibodies induced against human proteins expressed in bacteria and purified from inclusion bodies are not ideal for the identification of human proteins present on the cell membrane, as these quite often undergo post-translational modifications and structure conformations which are generally lost when the protein is expressed in bacteria. Consequently, monoclonal antibodies specific for human proteins expressed in bacteria have the limiting factor of the number of antibodies that need to be screened to find the ones that recognise the human proteins in human cells. We therefore utilized polyclonal antisera that include a combination of specificities in the same sample. On the other hand, making human antigens as his-tag proteins in bacteria has several practical advantages such as the higher throughput-working pipeline, the higher amount of proteins produced and the higher homogeneity of the different batches.

Since these type of antibodies quite often recognise on the native

form of human proteins only primary sequence structure, i.e., linear epitopes, and frequently they are even specific for epitopes that are not present on the "real life" proteins, it was important to obtain a proof of concept that our approach was suited to identify proteins expressed on human cell membranes. Therefore, we produced in bacteria, antisera specific for twenty well characterized proteins known to be present on hematopoietic cells and showed that we could identify 60% of known proteins on the surface of living PBLs by flow cytometry and 85% of known proteins in fixed and embedded lymphoid tissues by IHC. A likely explanation for the superiority of IHC versus flow cytometry with these antisera, relies on the nature of the antigens used to immunize mice. Indeed, human proteins on fixed and embedded cells, rather than on living cells, are likely to share more "denatured" epitopes with the same human proteins expressed in bacteria and purified as inclusion bodies. There were three main reasons that made this apparently inferior choice more suited for our purposes: 1) The easiest access to blood samples rather than lymph node biopsies. 2) The possibility to perform multiple colors staining on the cells of interest. 3) The possibility to gate and analyze fractions of very minor cells subsets (< 1-2%) which would pass mostly undetected in a screening performed by IHC.

Another feature of antibodies specific for linear epitopes is that, in

general, they display a lower avidity as compared with conformation-specific antibodies[25,26]. This decreases the possibility to use them to inhibit functions or transduce signals. Thus, this antibodies library shall be used only to identify a target protein, and later on monoclonal antibodies specific for the target proteins should be generated in order to perform functional studies.

Remarkably a large majority of the newly identified proteins are expressed on a fraction of B cells. This is somehow expected since a lower number of B cell markers have been characterized as compared to T cell markers. For a long period of time CD4+ T cells have been considered the "master" regulators of the immune[27]responses and a lot of functionally distinct T helper or regulatory subsets have been described and characterized[28-35]. B cells were generally considered antibody- producing effector cells and a combination of few surface markers was used to discriminate between human naïve B cells, memory B cells (central memory) and antibody-producing plasma cells (effector memory)[36]. However, B cells are more heterogeneous than previously thought. Although usually overshadowed by the production of antibodies, the ability of B cells to play important antibody-independent functions (antigen presentation, T cell and Dendritic cell regulation and cytokine and chemokine production) is well documented[37]. Trough these function B cells can profoundly influence the

formation and organization of secondary lymphoid tissues and T cell development, activation and function[37]. Moreover, antibody-independent B cell functions can contribute either to the development or to the prevention of autoimmune diseases[38]. It seems reasonable to assume that a larger amount of functionally distinct subsets variably contribute to the antibody-independent functions of B cells and that such subsets are defined by the expression of on or few neglected proteins, that we have identified with our polyclonal antisera library.

In the field of stem cells, the lack of surface markers, allowing the separation of stem cells subpopulations with a specific fate, represents the major problem in stem cell based therapies. Thus, the identification of such new stem cell markers would significantly improve their use in therapy. In the present study we aimed at identifying new hematopoietic stem cells subsets. Hematopoietic stem cell (HSC) transplantation is, nowadays, the only widely used stem cell-based therapy[39-41]. Even though HSC have been extensively studied in the last 20 years and a number of lineage-specific markers have been identified, the need of new subsets identification to better understand the hematopoiesis mechanisms is still strong. With our screening we have identified the protein MOSC-1 on a subset of HSC as well as on monocytes, where its expression was revealed also by gene expression profile. We

have performed a phenotypic characterization of MOSC-1 expressing HSC and we have purified by FACS Sorting the fraction of MOSC-1 positive cells, that generate in CFC assays almost only CFU-G and CFU-M colonies. These results strongly suggest that MOSC-1 is the marker defining the monocytes and granulocyte progenitors therefore identifying a functional HSC subset. We identified also the protein TREML-1 on a HSC subset. This protein is already known as a marker of trombin-activated platelets. Phenotypic analysis was performed for TREML-1 positive HSC, and these cells shown expression of markers that can indicate a possible role as megakaryocyte precursors. TREML-1 positive cells were then sorted by FACS to perform CFC, where the were able to generate all kind of colonies. Our hypothesis is they are very early megakaryocytes precursors, and has to be verified through differentiation assays.

In conclusion we generated a library of polyclonal antisera specific for unknown human surface proteins. This library is a powerful discovery tool. In fact our sera allow not only to identify new molecules expressed on a cell subset, but also to perform a phenotypic and functional pre-characterization of the newly identified cell subset. Moreover we were able to discover the expression of new proteins, not previously outlined with other methods, on HSC indicating that we have developed a very

sensitive approach particularly appropriate when working with poorly represented cells.

Materials and Methods

Preparation of a library of polyclonal antisera specific for human unknown proteins

The library of antisera used in the present work was produced as described in Supplementary material. Briefly the genes coding for proteins predicted to be transmembrane or secreted with unknown distribution and functions were selected and expressed in *E. coli*.

The selected genes were cloned in expression vectors under an inducible lac promoter and in frame with a 10 Histidine tag. This library was then transformed in E coli, and sequenced for validation.

The recombinant proteins were then purified by Protein Affinity Chromatography and used to immunized groups of five mice to produce polyclonal antisera.

Cell Preparation and Purification.

The PBL from healthy donors and cord blood were obtained by density gradient centrifugation on Lympholyte-H (Cedarlane Laboratories Ltd) and immediately analyzed after the separation. PBL activation was induced on overnight by the addition on medium phytohemagglutinin (PHA) 1 mg/ml (ROCHE Diagnostic

GmbH) and IL-2 100 U/ml (Novartis). The cells were cultured at 37 °C and 5% CO_2 in RMPI supplemented with 10%FBS (EuroClone S.p.A) and antibiotics (GIBCO)

Flow Cytometry.

Flow cytometry screening of the library was performed on 2 to 5 $X10^5$ PBL or $5x10^6$HSC. PBMCs resting or activated were stained in a three-step procedure. Incubating cells for 20 min at room temperature with 50% NHS (Euroclone) in PBS (Euroclone) to block Fc receptor; incubating with anti-serum at the optimal concentration (1:50 to 1:450 dilution) in 5%NHS PBS (FACS wash) for 10 min at 4°C, washing two time at 1500 rpm for 3 min and incubating with Goat anti mouse RPE (SouthernBiotech) at 1:200 dilution for 10 min at 4° C, washing two time with FACS wash at 1500 rpm for 3 min.

Multi colour FACS analysis was performed as following. Cells were stained for the first three steps as PBL then incubating with mouse IgG (SIGMA) at 4 mg/10^6 cells for 1 hour at 4°C then adding the conjugated mAbs: CD34, (IOT Coulter), CD45 (ImmunoTools), or for PBL in multi colour with CD3, CD19, CD56, CD71, Glycophorin A, CD7, CD33, CD38, CD10 (BD Biosciences), CD117 (IOT Coulter), CD133/2 (Macs Miltenyi Biotec GmbH), Mouse isotype-matched was used as negative controls. The samples were acquired using a FACS Canto II analyzer (Becton Dickinson) and data were processed by using the program FlowJo (Flow

Cytometry Analysis Software).

Amplification by RT PCR of the transcripts corresponding of the newly identified proteins

RNA extraction was performed using RNeasy Mini Kit or RNeasy Micro Kit (QIAGEN) on peripheral blood or cord blood cells lysates homogenized with QIAshredder homogenizer (QIAGEN).

cDNA synthesis was performed using SuperScript III First-Strand Synthesis SuperMix (Invitrogen).

RT-PCR primers 5'-3' sequences:

TMCC1	379 fw CAGGAGGAGCGATATAGATGTG
	379 rev TGGCTACAGTGGAGACAAAG
MOSC-1	194 fw TTCCTGAAGTCACAGCCCTAC
	194 rev GCATCTGGAACAAGCCATCAC
SUSD3	452 fw A TTGTGAGCTGTGCCATCATCC
	452 rev A TGTGGTGAAGCTGTGGTTGTC
TMEM126	314 fw GGCGACATTTGGAACAAC
	314 rev TTTGGTGGCAGTGGAACG
LPPR2	1174 fw AGCGATGTACGTGACTCTC
	1174 rev CAGTTCTGCGACTTGGATG
GSG1-L	444 fw b TCTGTCACCACGCTCAACTCC
	444 rev b AAGACCCAGCACTGTCGGTTC
TMEM38	147 fw CACCCAGCATCTGGCAATATC
	147 rev GCAACATCTACCGGCTTTGAG

KRTCAP3	665 fw AGGACTGCTGGATCCTCTG
	665 rev GCACCTGCTGTCCTAAACC
CRISP-1	26nested2 fw TAAGCTCGTCACCGACTTG
	26nested2 fw CTCCTCATCGTCACAGCATAG
	26a fw ACACAACGCCCTCAGGAGAAG
	26a rev TGGCGGCAAGATGCAATGG
	26b fw GTTTGGGCCACATCTTAC
	26b rev CGTCACAGCATAGAACAG

Immunohistochemistry

Sections of Human lymph nodes were pre-treated with an antigen retrieval solution and were then incubated with the indicated antisera at 1:100 dilution points. Detection steps were done using a commercially available kit according to the manufacturer's instruction. Peroxidase activity was developed with 3-3-diaminobenzidine-copper sulphate to obtain a brown-black end product.

Clonogenic Assay

Clonogenic assays were performed using αserum positive cells, isolated using FACS Aria cell sorter (Becton Dikinson). The cells were plated at 250 cells/plate in METHOCULT H4433 complete methylcellulose medium (StemCell Technologies). This formulation contains phytohemagglutinin-leukocyte conditioned medium (PHA-

LCM) as a source of colony stimulating factors, plus recombinant human erythropoietin. Cultures were incubated for 14 days in incubator adjusted to 37°C, 5%CO_2 and >95% humidity.

Transient Transfection

To carry out transient transfection experiments, we used MicroPorator MP-100, a pipette-type electroporation system (NanoEnTek Inc). The cells were dissociated by a brief treatment with trypsin-EDTA (Euroclone). Indicated plasmid DNAs were introduced into each $5X10^5$ dissociated cells in 10 μl volume according to manufacturer's instructions (3 pulses with 10 msec duration at 1600 voltages; Digital Bio Technology).

Electroporated cells were then seeded into 6-well culture dishes (Nunc) containing 2 ml of culture media. After 24-48 hrs of recovery, the cells were subjected to western blot analysis.

Western blot analysis

Whole-cell extracts were loaded on pre-casted sodium dodecyl sulfate (SDS)-polyacrylamide gels (Bio-Rad Laboratories S.r.l.) and transferred to Hybond-P PVDF membranes (GE Healthcare) using Towbin's buffer (25 mM Tris, pH 8.3, 192 mM glycine and 20% methanol). The blots were blocked in PBS containing 0.5% Tween 20 (SIGMA) and 5% not fat milk and incubated with primary antibody (anti-myc mAb 9E10 clone or antiserum) at room temperature for 1 hr. The blots were then washed four times with

PBS/0.5% Tween-20. Primary antibody binding was subsequently detected by incubation for 1 hour with secondary antibodies Goat anti mouse HRP (SouthernBiotech). The blots were then washed four times as described and were developed using the SuperSignal West Dura Extended Duration Substrate (Thermo Fisher Scientific)

References

1 Lanzavecchia, A., Maccario, R., Vitiello, A., Nespoli, L. & Ugazio, A. G. Human T lymphocyte subset identification according to avidity for sheep erythrocytes and rosetting capacity with allogeneic red blood cells. *Boll Ist Sieroter Milan* 56, 217-221 (1977).

2 Dieckmann, D., Plottner, H., Berchtold, S., Berger, T. & Schuler, G. Ex vivo isolation and characterization of CD4(+)CD25(+) T cells with regulatory properties from human blood. *The Journal of experimental medicine* 193, 1303-1310 (2001).

3 Jonuleit, H. *et al.* Identification and functional characterization of human CD4(+)CD25(+) T cells with regulatory properties isolated from peripheral blood. *The Journal of experimental medicine* 193, 1285-1294 (2001).

4 Langrish, C. L. *et al.* IL-23 drives a pathogenic T cell population that induces autoimmune inflammation. *The Journal of experimental medicine* 201, 233-240 (2005).

5 Annunziato, F. *et al.* Phenotypic and functional features of human Th17 cells. *The Journal of experimental medicine* 204, 1849-1861 (2007).

6 Hocquette, J. F. Where are we in genomics? *J Physiol Pharmacol* 56 Suppl 3, 37-70 (2005).

7 Haberkorn, U., Altmann, A. & Eisenhut, M. Functional genomics and proteomics--the role of nuclear medicine. *European journal of nuclear medicine and molecular imaging* 29, 115-132 (2002).

8 Loyet, K. M., Ouyang, W., Eaton, D. L. & Stults, J. T. Proteomic profiling of surface proteins on Th1 and Th2 cells. *Journal of proteome research* 4, 400-409 (2005).

9 Liu, F. *et al.* Insights into human CD34+ hematopoietic stem/progenitor cells through a systematically proteomic survey coupled with transcriptome. *Proteomics* 6, 2673-2692 (2006).

10 Foster, L. J. *et al.* Differential expression profiling of membrane proteins by quantitative proteomics in a human mesenchymal stem cell line undergoing osteoblast differentiation. *Stem cells (Dayton, Ohio)* 23, 1367-1377 (2005).

11 Hemmoranta, H. *et al.* Transcriptional profiling reflects shared and unique characters for CD34+ and CD133+ cells. *Stem cells and development* 15, 839-851 (2006).

12 Jaatinen, T. *et al.* Global gene expression profile of human cord blood-derived CD133+ cells. *Stem cells (Dayton, Ohio)* 24, 631-641 (2006).

13 Oh, J. M. *et al.* A database of protein expression in lung cancer. *Proteomics* 1, 1303-1319 (2001).

14 Xu, C. *et al.* Identification and characterization of 177 unreported genes associated with liver regeneration. *Genomics Proteomics Bioinformatics* 2, 109-118 (2004).

15 Madras, N., Gibbs, A. L., Zhou, Y., Zandstra, P. W. & Aubin, J. E. Modeling stem cell development by retrospective analysis of gene expression profiles in single progenitor-derived colonies. *Stem cells (Dayton, Ohio)* 20, 230-240 (2002).

16 Nierhoff, D. *et al.* New cell surface markers for murine fetal hepatic stem cells identified through high density complementary DNA microarrays. *Hepatology (Baltimore, Md* (2007).

17 Lyman, G. H. & Kuderer, N. M. Gene expression profile assays as predictors of recurrence-free survival in early-stage breast cancer: a metaanalysis. *Clin Breast Cancer* 7, 372-379 (2006).

18 Fischer, D. Servers for protein structure prediction. *Current opinion in structural biology* 16, 178-182 (2006).

19 Anantharaman, V. & Aravind, L. MOSC domains: ancient, predicted sulfur-carrier domains, present in diverse metal-sulfur cluster biosynthesis proteins including Molybdenum cofactor sulfurases. *FEMS microbiology letters* 207, 55-61 (2002).

20 Washington, A. V., Quigley, L. & McVicar, D. W. Initial characterization of TREM-like transcript (TLT)-1: a putative inhibitory receptor within the TREM cluster. *Blood* 100, 3822-3824, doi:10.1182/blood-2002-02-0523 2002-02-0523 [pii] (2002).

21 Allcock, R. J., Barrow, A. D., Forbes, S., Beck, S. & Trowsdale, J. The human TREM gene cluster at 6p21.1 encodes both activating and inhibitory single IgV domain receptors and includes NKp44. *European journal of immunology* 33, 567-577, doi:10.1002/immu.200310033 (2003).

22 Barrow, A. D. *et al.* Cutting edge: TREM-like transcript-1, a platelet immunoreceptor tyrosine-based inhibition motif encoding costimulatory immunoreceptor that enhances, rather than inhibits, calcium signaling via SHP-2. *J Immunol* 172, 5838-5842 (2004).

23 Washington, A. V. *et al.* A TREM family member, TLT-1, is found exclusively in the alpha-granules of megakaryocytes and platelets. *Blood* 104, 1042-1047, doi:10.1182/blood-2004-01-0315 2004-01-0315 [pii] (2004).

24 Gattis, J. L. *et al.* The structure of the extracellular domain of triggering receptor expressed on myeloid cells like transcript-1 and evidence for a naturally occurring soluble fragment. *The Journal of biological chemistry* 281, 13396-13403, doi:M600489200 [pii] 10.1074/jbc.M600489200 (2006).

25 Ando, T., Latif, R., Daniel, S., Eguchi, K. & Davies, T. F. Dissecting linear and conformational epitopes on the native thyrotropin receptor. *Endocrinology* 145, 5185-5193, doi:10.1210/en.2004-0789 en.2004-0789 [pii] (2004).

26 Hijnen, M. *et al.* Epitope structure of the Bordetella pertussis protein P.69 pertactin, a major vaccine component and protective antigen. *Infection and immunity* 72, 3716-3723, doi:10.1128/IAI.72.7.3716-3723.2004 72/7/3716 [pii] (2004).

27 Lemoine, S., Morva, A., Youinou, P. & Jamin, C. Regulatory B cells in autoimmune diseases: how do they work? *Annals of the New York Academy of Sciences* 1173, 260-267, doi:NYAS4651 [pii] 10.1111/j.1749-6632.2009.04651.x (2009).

28 Bluestone, J. A., Mackay, C. R., O'Shea, J. J. & Stockinger, B. The functional plasticity of T cell subsets. *Nature reviews* 9, 811-816, doi:nri2654 [pii] 10.1038/nri2654 (2009).

29 Turtle, C. J., Swanson, H. M., Fujii, N., Estey, E. H. & Riddell, S. R. A distinct subset of self-renewing human memory CD8+ T cells survives cytotoxic chemotherapy. *Immunity* 31, 834-844, doi:S1074-7613(09)00452-X [pii]

10.1016/j.immuni.2009.09.015 (2009).
30 Eyerich, S. *et al.* Th22 cells represent a distinct human T cell subset involved in epidermal immunity and remodeling. *The Journal of clinical investigation* 119, 3573-3585, doi:40202 [pii] 10.1172/JCI40202 (2009).
31 Weaver, C. T. & Hatton, R. D. Interplay between the TH17 and TReg cell lineages: a (co-)evolutionary perspective. *Nature reviews* 9, 883-889, doi:nri2660 [pii] 10.1038/nri2660 (2009).
32 Annunziato, F. & Romagnani, S. Heterogeneity of human effector CD4+ T cells. *Arthritis research & therapy* 11, 257, doi:ar2843 [pii] 10.1186/ar2843 (2009).
33 Chan, W. L. *et al.* NKT cell subsets in infection and inflammation. *Immunology letters* 85, 159-163, doi:S0165247802002237 [pii] (2003).
34 Harari, A. *et al.* Functional signatures of protective antiviral T-cell immunity in human virus infections. *Immunological reviews* 211, 236-254, doi:IMR395 [pii] 10.1111/j.0105-2896.2006.00395.x (2006).
35 Kallies, A. Distinct regulation of effector and memory T-cell differentiation. *Immunology and cell biology* 86, 325-332, doi:icb200816 [pii] 10.1038/icb.2008.16 (2008).
36 Dogan, I. *et al.* Multiple layers of B cell memory with different effector functions. *Nature immunology* 10, 1292-1299, doi:ni.1814 [pii] 10.1038/ni.1814 (2009).
37 Sanz, I., Wei, C., Lee, F. E. & Anolik, J. Phenotypic and functional heterogeneity of human memory B cells. *Seminars in immunology* 20, 67-82, doi:S1044-5323(07)00118-2 [pii] 10.1016/j.smim.2007.12.006 (2008).
38 Dorner, T., Jacobi, A. M. & Lipsky, P. E. B cells in autoimmunity. *Arthritis research & therapy* 11, 247, doi:ar2780 [pii] 10.1186/ar2780 (2009).
39 Gratwohl, A. *et al.* Current trends in hematopoietic stem cell transplantation in Europe. *Blood* 100, 2374-2386 (2002).

40 Gluckman, E. & Rocha, V. History of the clinical use of umbilical cord blood hematopoietic cells. *Cytotherapy* 7, 219-227 (2005).
41 Bach, F. H., Albertini, R. J., Joo, P., Anderson, J. L. & Bortin, M. M. Bone-marrow transplantation in a patient with the Wiskott-Aldrich syndrome. *Lancet* 2, 1364-1366 (1968).

Chapter 3

Use of the newly identified stem cell markers to better characterize the hematological and immunological reconstitution after HSCT

Ongoing work

A population of immature cells are present in PBLs of HSCT patients at the time of the engraftment.

Three different sources of hematopoietic stem cells are commonly used for transplantation of pediatric patients: bone marrow (BM), cord blood (UCB) and haploidentical hematopoietic stem cells from mobilized peripheral blood (MPB-SC). To focus on the early dynamics of immune system-reconstitution in the presence or in the absence of immune-suppression we perform an immunophenotypic analysis of PBMC from HSCT children underwent BMT, UCBT or MPB-SCT at the time of the engraftment. We found that while in a healthy subject the population of peripheral blood lymphocytes (PBL) is almost completely composed by CD3+ (T cells), CD19+ (B cells) and CD56+ (NK) cells (**Figure 1A**), there is a consistent amount of cells that lack the expression of the "mature" lineage markers CD3, CD19 and CD56 within the PBLs of all the HSCT patients analyzed (**Figure 1B**). This population of Immature Circulating Cells (ICC) represent a large proportion of patients PBLs at the time of the engraftment and decreases progressively in the months thereafter to reach the same percentage found in healthy children (**Figure 1C**). To assess whether ICC percentage at the time of the engraftment is related to the source of HSC used, we have grouped the patients according the HSCT received and we have represented the ratio between mature lymphocytes and ICC at the

time of the engraftment. As shown in **Figure 1D** there are no significant differences in the percentage of ICC in the different groups of patients indicating that the source of HSC used for the transplantation does not affect the presence of ICC at the time of the engraftment.

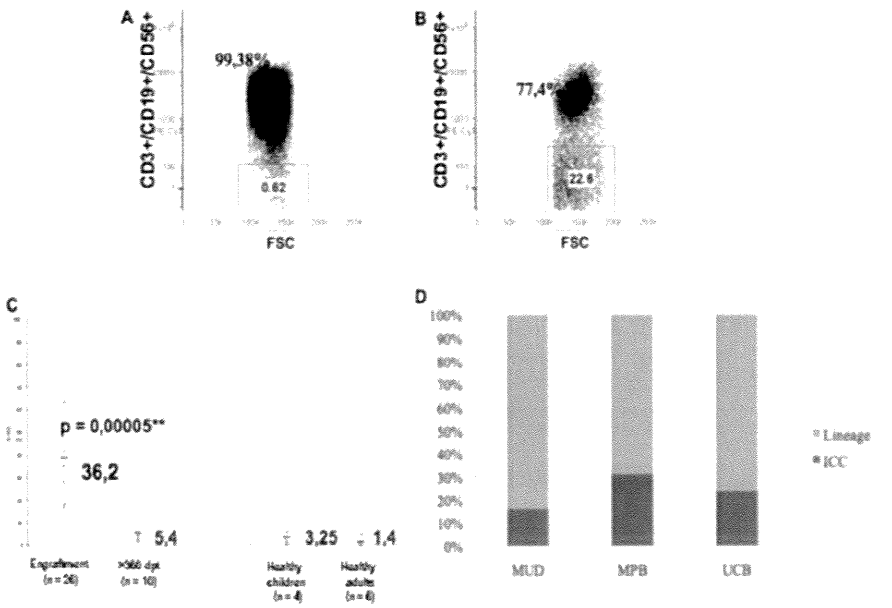

Figure 1. A non-conventional population is over-represented in the peripheral blood lymphocytes of children after HSCT.

PBMCs are obtained by Ficoll separation from peripheral blood of healthy children (median age 4y, 2-16y) and HSCT patients (median age 4y9m, 10m-16y). The cells are stained with the indicated Pe-Cy7-conjugated antibodies and TOPRO-3 to exclude death cells. The analysis within the region of living lymphocytes is shown.
A) Expression of CD3, CD19 and CD56 in healthy children PBLs. A representative plot is shown. The percentage of cells with immature phenotype (CD3 neg., CD19 neg., CD56 neg.,) is indicated in the square box.
B) Expression of CD3, CD19 and CD56 in PBLs of HSCT patients at the engraftment. A representative plot is shown. The percentage of cells with immature phenotype (CD3 neg., CD19 neg., CD56 neg.,) is indicated in the square box.
C) ICC population is poorly represented at late time points after HSCT. Comparison of ICC percentage in HSCT patients at the time of the engraftment and HSCT patients >360 days post transplantation (dpt). ICCs are identified by FACS on PBMCs as CD3 negative, CD19 negative, CD56 negative cells within the region of living lymphocytes. The percentage of ICC subset in healthy children and healthy adults is reported.
D) ICCs versus mature cells distribution in children receiving HSCT from different sources.

ICCs are variously composed by hematological precursors.

Since ICCs have not been described yet, we performed a phenotypic and functional characterization of this subset. Assuming that ICCs subset was likely composed by different hematological precursors, we have chosen different combination of markers allowing to discriminate between hematopoietic stem cells or progenitor cells (CD34+CD45dim, CD133+), early lymphoid precursors (CD38+), lymphoid precursors (CD45RA +, CD7+), T cells precursors (CD45RA +, CD7+, CD2+, CD1a+), NK precursors (CD45RA+, CD7+, CD161+, CD122+/-, CD94+/-), myeloid precursors (CD123+), eritroid precursors (CD71+, CD45+/-, CD235a+/-). We performed the phenotypic analysis on fresh blood samples to minimize any change in the phenotype of the cells ex vivo. A combination of "mature lineage" markers (CD3, CD19, CD56, CD14) was always used to exclude all the mature cells from the analysis. In **Figure 2** is shown how the various hematological precursors contribute to the ICC composition in patients receiving HSC from the different sources. For each subset it is reported the average of all the patients analyzed. Our data showed that eritroid precursors are significantly higher in patients receiving CBT (Cord Blood Transplantation) as compared to patients receiving HSC from other sources, and that the phenotype of a consistent ICC percentage remains completely elusive (**Figure 2**).

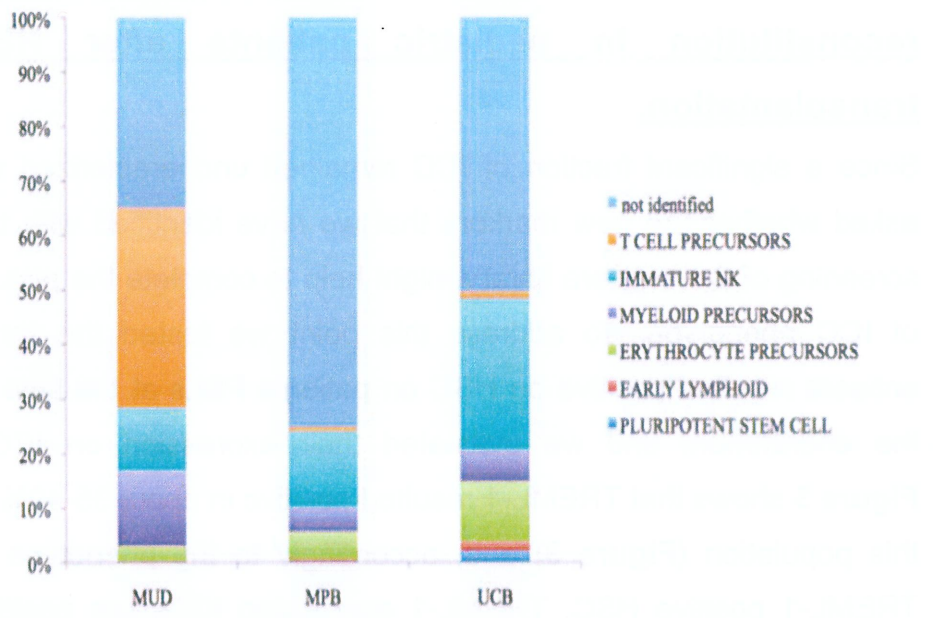

Figure 2. ICC phenotype.
Patients were sorted according with the source of HSC for the transplantation. Their blood was analyzed by flow cytometry. The analysis is performed on the gate of CD3 negative, CD19 negative, CD56 negative living lymphocytes. In the graph is reported the average of the precursors subpopulations in the different groups of patients.
Cells negative for such markers were considered ICC and stained to identify the following hematological precursor subsets: T cells precursors (CD45RA +, CD7+, CD2+, CD1a+), Immature NK (CD45RA +, CD7+, CD161+, CD122+/-, CD94+/-), Myeloid Precursors (CD123+), Eritrocyte precursors (CD71+, CD45+/-, CD235a+/-), Early Lymphoid Precursors (CD45RA +, CD7+), pluripotent HSC or progenitor cells (CD34+, CD45dim, CD133+). A large ICC percentage resulted not classifiable in these subsets and we considered it "not identified".
(MUD = Matched Unrelated Donor; MPB = Mobilized Peripheral Blood from an haploidentical donor; UCB = Umbilical Cord Blood).

TREML-1 as a potential marker of platelet reconstitution in pediatric patients after HSC transplantation.

Since a significant fraction of ICC remained uncharacterized we asked whether the new markers that we have identified with the screening of the antisera library might help to complete the picture of ICC phenotype. To address this point we tested the three antisera resulted positive on HSC on patients PBLs at the time of the engraftment and we evaluated their expression on ICCs. **Figure 3** shows that TREML-1 resulted positive in about 35-40% of this population (**Figure 3**) and, accordingly to the phenotype of TREML-1 positive HSC, TREML-1 expressing ICCs are positive also for the expression of CD33 (**Figure 3C**) These data indicate that anti TREML-1 antibodies could be used to identify megakariocytes precursors in peripheral blood of HSCT patients and, therefore, the protein could be used as marker of platelet reconstitution.

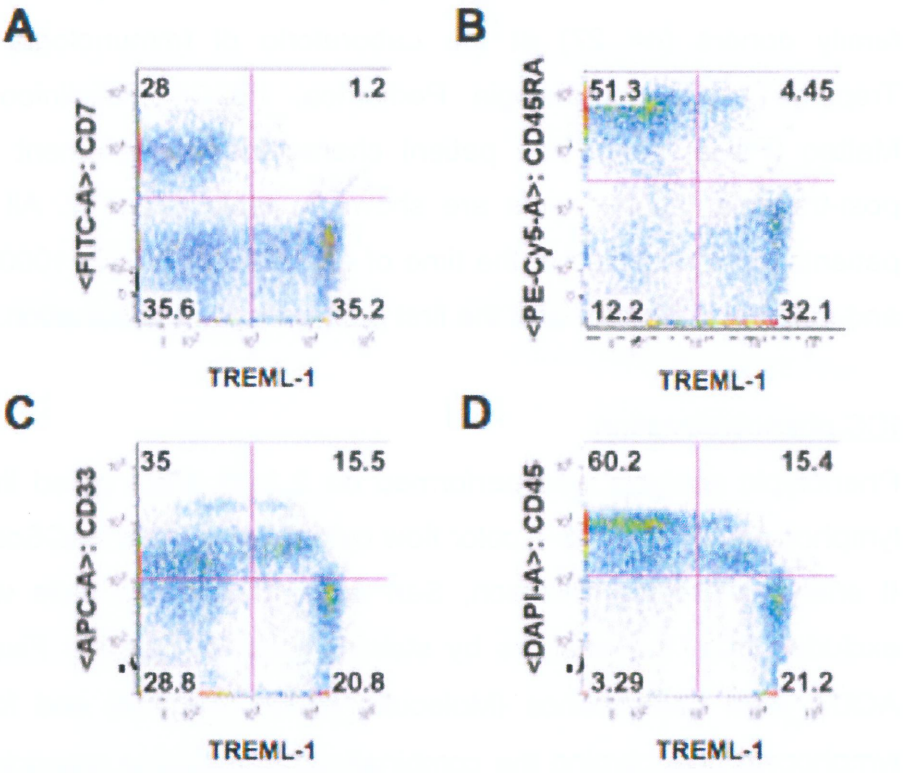

Figure 3. TREML-1 on ICCs.
Flow cytometry analysis of TREML-1 presence on ICC from a pediatric patient (plots are representative of 3 independent experiments). PBMC from patient were stained with the "exclusion mix" (anti CD3, anti CD14, anti CD19 and anti CD56). Cells for these markers in the PBLs' region are considered ICCs. ICCs were stained with TREML-1 in combination with anti CD7, anti CD45RA, anti CD33 and anti CD45. TREML-1 positive cells were about 36%. They expressed little or not CD7 and CD45RA (A and B), while little less than half of them shown presence of CD33 (C) and CD45 (D).

Materials and Methods

Patients' characteristic and blood samples

We analyzed N patients treated with CBT (n = 16), BMT from matched unrelated donors (n = 5) or MPB from haploidentical family donors (n= 27) at the Laboratorio di Immunologia dei Trapianti e Oncoematologia Pediatrica, IRCCS Policlinico S. Matteo (Pavia, Italy). The patient characteristics, treatment and post-transplantation events are shown in table 1 and 2. All the patients were analyzed at the time of engraftment (WBC>1000/ml) and monthly thereafter until the first post-transplant vaccination.

ICC phenotypizzation

Phenotypic analysis was performed on 2 to 5 X104 gated living lymphocytes by direct 5-7 color flow cytometry using a FACScanto II analyzer (Becton Dikinson, San Jose, CA). Dead cells were excluded from the analysis by staining with LIVE/DEAD Fixable AQUA stain fluorescence (Molecular Probe, Eugene) and for T lymphocyte phenotyping the combinations of following monoclonal antibodies (mAbs) and appropriate isotype-matched controls.

Mix 1: Ki67 Fitc (BD San Jose, CA)/CD3Pe (BD San Jose, CA)/CD19PCY5 (BD San Jose, CA)/CD56PCY7 (Immunotech SAS, France)/CD71 APC (BD San Jose, CA)/CD45Vio-Blu (MILTENYI BIOTEC, GERMANY). **Mix 2**: CD38 Pe (BD San Jose, CA)/Glycophorine A PCY5 (BD San Jose, CA)/CD71 APC (BD San Jose, CA)/CD45 APCY7(BD San Jose, CA)/CD3PC7 (Immunotech

SAS, France)/CD19PC7 (BD San Jose, CA)/CD56PC7 (Immunotech SAS, France)/CD14PC7 (Immunotech SAS, France). **Mix 3**: CD7 Fitc (BD San Jose, CA)/CD123Pe (BD San Jose, CA)/CD45RA PCY5 (BD San Jose, CA)/DR APC (BD San Jose, CA)/ CD2 Bio (BD San Jose, CA) Sav-APCY7 (BD San Jose, CA)/ CD3PC7 (Immunotech SAS, France)/CD19PC7 (BD San Jose, CA)/CD56PC7 (Immunotech SAS, France)/CD14PC7 (Immunotech SAS, France). **Mix4**: CD7 Fitc (BD San Jose, CA)/CD10 Pe (BD San Jose, CA)/CD45RA PCY5 (BD San Jose, CA)/CD33 APC (BD San Jose, CA)/CD123 Bio (BD San Jose, CA) Sav-APCY7 (BD San Jose, CA)/CD3PC7 (Immunotech SAS, France)/CD19PC7 (BD San Jose, CA)/CD56PC7 (Immunotech SAS, France)/CD14PC7 (Immunotech SAS, France). **Mix5:** CD7 Fitc (BD San Jose, CA)/CD38Pe (BD San Jose, CA)/CD45RA PCY5 (BD San Jose, CA)/CD33 APC (BD San Jose, CA)/CD123 Bio (BD San Jose, CA) Sav-APCY7 (BD San Jose, CA)/CD3PC7 (Immunotech SAS, France)/CD19PC7 (BD San Jose, CA)/CD56PC7 (Immunotech SAS, France)/CD14PC7 (Immunotech SAS, France). **Mix6:** CD7 Fitc (BD San Jose, CA)/CD127Pe (BD San Jose, CA)/CD45RA PCY5 (BD San Jose, CA)/CD161 APC (BD San Jose, CA)/ CD122 Bio (BD San Jose, CA) Sav-APCY7 (BD San Jose, CA)/CD3PC7 (Immunotech SAS, France)/CD19PC7 (BD San Jose, CA)/CD56PC7 (Immunotech SAS, France)/CD14PC7 (Immunotech SAS, France). **Mix7**: CD31 Fitc (BD San Jose, CA)/CD38 Pe (BD San Jose, CA)/CD34 PCY5

(Immunotech SAS, France)/CD133 APC (Miltenyi Biotec, Germany)/CD90 Bio (eBioscience) Sav-APCY7(BD San Jose, CA)/CD45 Vio-Blu (Miltenyi Biotec, Germany) / CD3PC7(Immunotech SAS, France)/CD19PC7 (BD San Jose, CA)/CD56PC7 (Immunotech SAS, France)/CD14PC7 (Immunotech SAS, France). **Mix8:** CD3 Fitc (BD San Jose, CA)/ CD62L Pe (BD San Jose, CA)/CD45RA PCY5 (BD San Jose, CA)/CCR7 PCY7 (BD San Jose, CA)/CD4 APCY-7(BD San Jose, CA).

Antisera was tested on these cells with the same multi color protocol used for CB cells (see *Cap.2 - Materials and Methods*), but in combination with the following mAb: aCD31 FITC, aCD61 PE (BD Pharmigen), aCD41 FITC, aCD 63 FITC, aCD29 APC (ImmunoTools), aCD90 PC5, aCD34 aPC7 (IOT Coulter), aCD133/2 APC, aCD45 Violet Blu (Miltenyi Biotec). The antiserum was tested 1:50, with a goat anti mouse RPE (SouthernBiotech), or a goat anti mouse AlxaFluor 633 (Invitrogen), according to the mAb mix.

Chapter 4

Summary, conclusions and future perspectives

We designed a new approach to identify cell subsets based on a library of mouse polyclonal antisera specific for linear epitopes of poorly known human proteins that were predicted to be either transmebrane or secreted, and we have shown that with this library it is possible to interrogate by flow cytometry the cell surface of hematopoietic cells and to identify new subsets of both mature and hematopoietic cells. This library is versatile—it can be used to screen any cell or tissue of interest—and allows screening of a large (1600) repertoire of "neglected" human proteins for those that mark specifically new subsets within apparently homogeneous cell populations. The workflow of this project is summarized in **Figure1**.

The identification of new proteins on cells or tissues is generally based either on transcriptomics, i.e., the assessment of mRNA expression profile, or on proteomics[1-6], i.e., the direct identification by mass spectrometry of proteins separated by 2D gels or liquid-based separation methods. Proteomics is very informative but poses the problem of the complexity of the approach that makes it not suitable for high throughput screenings. Transcriptomics allows analyzing and compare large amount of samples at the same time[7], but poses the problem of the correlation between mRNA levels and protein expression levels. Moreover, to perform

a RNA microarray analysis at least 50ng of high quality RNA is required, independently from the chosen platform. To reach this amount of RNA from stem cell populations is a difficult task: stem cells are not only poorly represented, but also quiescent[7-13]. Thus, the amount of RNA that can be extracted from stem cells is lower than what we can expect from a comparable number of differentiated cells of a given tissue. For this reason a gene, expressed from just a subset of stem cells, is difficult to detect with transcriptomics. Different methods of RNA pre-amplification have been developed to overcome this limitation[14-22]. However, these methods suffer the fact that the RNA is manipulated before the microarray analysis and there is always the possibility that the transcript representation is affected by the treatment leading to loose some signals while overrating others[23-27].

In both cases, it is impossible to assess whether differences in the expression levels of genes or proteins occurs in all of the cells analyzed or in a subset of them. It is therefore difficult to study those cell subsets or lineages that are poorly represented within a population. Our goal was to study hematopoietic cell subsets by flow cytometry and therefore opted for an approach based on the direct identification of proteins with an antibody library.

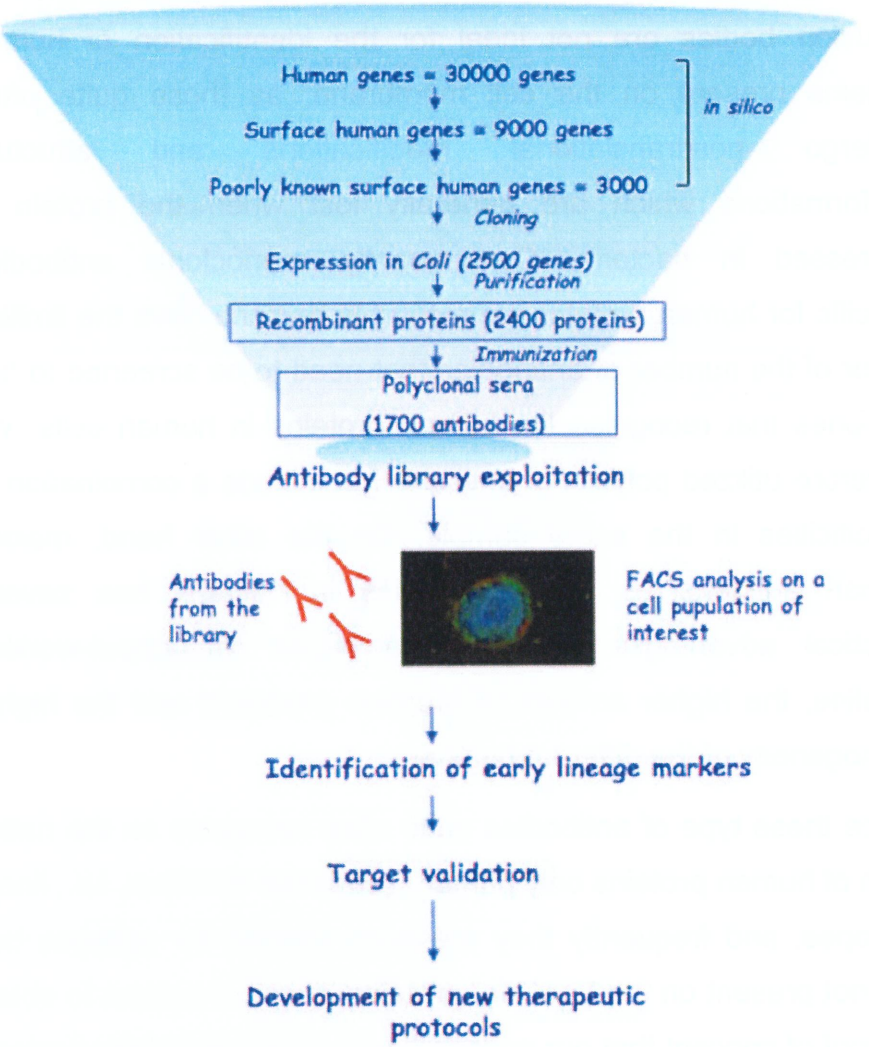

Figure 10. Summary and Future Perspectives.
The figure summarizes the entire project. At present time we are in the phase of target validation. Our results are very promising and we possibly will use thise new knowledge to develop new therapeutic tools.

We elicited our antibody library using proteins expressed in bacteria as immunogens. On the one hand, antibodies induced against human proteins expressed in bacteria and purified from inclusion bodies are not ideal for the identification of human proteins present on the cell membrane, as these quite often undergo posttranslational modifications and structure conformations which are generally lost when the protein is expressed in bacteria. Consequently, monoclonal antibodies specific for human proteins expressed in bacteria have the limiting factor of the number of antibodies that need to be screened to find the ones that recognise the human proteins in human cells. We therefore utilized polyclonal antisera that include a combination of specificities in the same sample. On the other hand, making human antigens as his-tag proteins in bacteria has several practical advantages such as the higher throughput-working pipeline, the higher amount of proteins produced and the higher homogeneity of the different batches.

Since these type of antibodies quite often recognise on the native form of human proteins only primary sequence structure, i.e., linear epitopes, and frequently they are even specific for epitopes that are not present on the “real life” proteins, it was improtant to obtain a proof of concept that our approach was suited to identify proteins expressed on human cell membranes. Therefore, we produced in bacteria, antisera specific for twenty well characterized proteins known to be present on hematopoietic cells and showed that we

could identify 60% of known proteins on the surface of living PBLs by flow cytometry and 85% of known proteins in fixed and embedded lymphoid tissues by IHC. A likely explanation for the superiority of IHC versus flow cytometry with these antisera relies on the nature of the antigens used to immunize mice. Indeed, human proteins on fixed and embedded cells, rather than on living cells, are likely to share more "denatured" epitopes with the same human proteins expressed in bacteria and purified as inclusion bodies. There were three main reasons that made this apparently inferior choice more suited for our purposes: 1) The easiest access to blood samples rather than lymph node biopsies. 2) The possibility to perform multiple color staining on the cells of interest. 3) The possibility to gate and analyze fractions of very minor cells subsets (< 1-2%) which would pass mostly undetected in a screening performed by IHC.

Another feature of antibodies specific for linear epitopes is that, in general, they display a lower avidity as compared with conformation-specific antibodies[20, 21] This decreases the possibility to use them to inhibit functions or transduce signals. Thus, this antibodies library shall be used only to identify a target protein, and later on monoclonal antibodies specific for the target proteins should be generated in order to perform functional studies.

In the field of stem cells, the lack of surface markers, allowing the separation of stem cells subpopulations with a specific fate,

represents the major problem in stem cell based therapies. Thus, the identification of such new stem cell markers would significantly improve their use in therapy. In the present study we aimed at identifying new hematopoietic stem cells subsets. Hematopoietic stem cell (HSC) transplantation is, nowadays, the only widely used stem cell-based therapy[34-36]. Even though HSC have been extensively studied in the last 20 years and a number of lineage-specific markers have been identified, the need of new subsets identification to better understand the hematopoiesis mechanisms is still strong. With our screening we have identified three new proteins expressed by a subset of HSC. Interestingly two thereof these proteins (CRISP-1 and MOSC-1) were not previously outlined by classical molecular analysis.

We have performed a phenotypic characterization of MOSC-1 expressing HSC and we have purified by FACS Sorting the fraction of MOSC-1 positive cells, that generate in CFC assays almost only CFU-G and CFU-M colonies. These results strongly suggest that MOSC-1 is the marker defining the monocytes and granulocyte progenitors. This finding may have important therapeutic applications. In fact it is possible, at least in principle, to use MOSC-1 to either delete malignant cells of myeloid origin or to enrich a graft for transplantation with monocytes progenitors. Another protein that we found expressed on a fraction of HSC is

TREM-1, a trombin-activated platelet marker. Again, we characterized phenotypically cells positive for this protein, found many of them cells expressing the platelet makers CD41 and CD61, and sorted them to perform CFC assays, were they gave origin to all possible hematopoietic colonies. With these results we concluded TREML-1 positive cells are quite immature progenitors that can still differentiate in several kinds of hematopoietic effector cells, but already addressed into megakaryocyte commitment. This protein resulted expressed also in a small fraction of a newly identified immature circulating population in HSC-transplanted children, with the same potential significance. The development potential of TREML-1 cells has now to be firmly confirmed by differentiation assays we are starting on CB HSCs. With the support of this evidence, we could claim TREML-1 positive HSCs should be very early megakaryocyte precursors, and they could possibly become very useful in the definition of engraftment prognosis after HSC.

In conclusion we generated a library of polyclonal antisera specific for unknown human surface proteins. This library is a powerful discovery tool. In fact our antisera allow not only to identify new molecules expressed on a cell subset, but also to perform a phenotypic and functional pre-characterization of the newly identified cell subset. Moreover we were able to discover the

expression of new proteins, not previously outlined with other methods, on HSC or PBMC, indicating that we have developed a very sensitive approach particularly appropriate when working with poorly represented cells. The complete functional characterization of the populations identified in our studies is the logical extension of the project.

We have started the validation phase of the newly identified targets by increasing the functional and differentiation studies. From this point of view, we have also started collaboration with other institutions to extend the possible studies to perform. For example, in collaboration with prof. Tortora of Bicocca University, Milan, we are performing studies of protein-protein interaction on SUSD3 protein, one of the markers we have identified on B-lymphocytes.

In the next future our antisera library will be screened on different cell population. We already performed the first step of the screening procedure on human BM-derived mesenchymal stem cells, in collaboration with S. Matteo's Pediatric Oncohematology unit. On MSC two sera resulted positive after the first screening and we are starting the second step.

Finally, a key issue for the advancement of this project is the generation of monoclonal antibodies against the protein we intend to characterize in detail. To do that we are starting to produce the proteins we have identified in mammalian cells in order to get native proteins for mice immunization and generation of hybridomas.

References

1 Loyet, K. M., Ouyang, W., Eaton, D. L. & Stults, J. T. Proteomic profiling of surface proteins on Th1 and Th2 cells. *Journal of proteome research* **4**, 400-409, doi:10.1021/pr049810q (2005).

2 Liu, F. *et al.* Insights into human CD34+ hematopoietic stem/progenitor cells through a systematically proteomic survey coupled with transcriptome. *Proteomics* **6**, 2673-2692 (2006).

3 Foster, L. J. *et al.* Differential expression profiling of membrane proteins by quantitative proteomics in a human mesenchymal stem cell line undergoing osteoblast differentiation. *Stem cells (Dayton, Ohio)* **23**, 1367-1377 (2005).

4 Hemmoranta, H. *et al.* Transcriptional profiling reflects shared and unique characters for CD34+ and CD133+ cells. *Stem cells and development* **15**, 839-851 (2006).

5 Jaatinen, T. *et al.* Global gene expression profile of human cord blood-derived CD133+ cells. *Stem cells (Dayton, Ohio)* **24**, 631-641 (2006).

6 Wagner, W. *et al.* The heterogeneity of human mesenchymal stem cell preparations--evidence from simultaneous analysis of proteomes and transcriptomes. *Experimental hematology* **34**, 536-548 (2006).

7 Yoshikawa, K. Cell cycle regulators in neural stem cells and postmitotic neurons. *Neuroscience research* **37**, 1-14 (2000).

8 Fortunel, N. *et al.* Release from quiescence of primitive human hematopoietic stem/progenitor cells by blocking their cell-surface TGF-beta type II receptor in a short-term in vitro assay. *Stem cells (Dayton, Ohio)* **18**, 102-111 (2000).

9 Ezoe, S., Matsumura, I., Satoh, Y., Tanaka, H. & Kanakura, Y. Cell cycle regulation in hematopoietic stem/progenitor cells. *Cell cycle (Georgetown, Tex* **3**, 314-318 (2004).

10 Walkley, C. R., Fero, M. L., Chien, W. M., Purton, L. E. & McArthur, G. A. Negative cell-cycle regulators cooperatively control self-renewal and differentiation of haematopoietic stem cells. *Nature cell biology* **7**, 172-178 (2005).

11 Larsson, J., Blank, U., Klintman, J., Magnusson, M. & Karlsson, S. Quiescence of hematopoietic stem cells and maintenance of the

stem cell pool is not dependent on TGF-beta signaling in vivo. *Experimental hematology* **33**, 592-596 (2005).

12 Suda, T., Arai, F. & Hirao, A. Hematopoietic stem cells and their niche. *Trends in immunology* **26**, 426-433 (2005).

13 Arai, F. & Suda, T. Maintenance of quiescent hematopoietic stem cells in the osteoblastic niche. *Annals of the New York Academy of Sciences* **1106**, 41-53 (2007).

14 Schuler, S., Wenz, I., Wiederanders, B., Slickers, P. & Ehricht, R. An alternative method to amplify RNA without loss of signal conservation for expression analysis with a proteinase DNA microarray in the ArrayTube format. *BMC genomics* **7**, 144 (2006).

15 Nygaard, V. & Hovig, E. Options available for profiling small samples: a review of sample amplification technology when combined with microarray profiling. *Nucleic acids research* **34**, 996-1014 (2006).

16 Ginsberg, S. D. RNA amplification strategies for small sample populations. *Methods (San Diego, Calif* **37**, 229-237 (2005).

17 Singh, R. *et al.* Microarray-based comparison of three amplification methods for nanogram amounts of total RNA. *Am J Physiol Cell Physiol* **288**, C1179-1189 (2005).

18 Stirewalt, D. L. *et al.* Single-stranded linear amplification protocol results in reproducible and reliable microarray data from nanogram amounts of starting RNA. *Genomics* **83**, 321-331 (2004).

19 Li, Y. *et al.* Systematic comparison of the fidelity of aRNA, mRNA and T-RNA on gene expression profiling using cDNA microarray. *Journal of biotechnology* **107**, 19-28 (2004).

20 Xiang, C. C. *et al.* A new strategy to amplify degraded RNA from small tissue samples for microarray studies. *Nucleic acids research* **31**, e53 (2003).

21 Wang, D. *et al.* Microarray-based detection and genotyping of viral pathogens. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 15687-15692 (2002).

22 Iscove, N. N. *et al.* Representation is faithfully preserved in global cDNA amplified exponentially from sub-picogram quantities of mRNA. *Nature biotechnology* **20**, 940-943 (2002).

23 Jung, S. H. & Jang, W. How accurately can we control the FDR in analyzing microarray data? *Bioinformatics (Oxford, England)* **22**, 1730-1736 (2006).

24 Subkhankulova T, L. F. Comparative evaluation of linear and exponential amplification techniques for expression profiling at the single-cell level. *Genome biology* **7** (2006).

25 Wick, N. *et al.* Nonuniform hybridization: a potential source of error in oligonucleotide-chip experiments with low amounts of starting material. *Diagn Mol Pathol* **13**, 151-159 (2004).

26 van Haaften, R. I. *et al.* Biologically relevant effects of mRNA amplification on gene expression profiles. *BMC bioinformatics* **7**, 200 (2006).

27 Degrelle, S. A. *et al.* Amplification biases: possible differences among deviating gene expressions. *BMC genomics* **9**, 46 (2008).

28 Gluckman, E. & Rocha, V. History of the clinical use of umbilical cord blood hematopoietic cells. *Cytotherapy* **7**, 219-227 (2005).

29 Gratwohl, A. *et al.* Current trends in hematopoietic stem cell transplantation in Europe. *Blood* **100**, 2374-2386 (2002).

30 Bach, F. H., Albertini, R. J., Joo, P., Anderson, J. L. & Bortin, M. M. Bone-marrow transplantation in a patient with the Wiskott-Aldrich syndrome. *Lancet* **2**, 1364-1366 (1968).

Appendix

Antisera List.

EntrezGeneID	Accession Transcript	Symbol	Length_AA	Nterm_AA	Cterm_AA
1	BC035719	A1BG	93	85	177
1	BC011406	A1BG	340	1	340
167	NM_170609	CRISP1	249	1	249
416	BC022369	ARSF	356	235	590
419	BC017913	ART3	354	1	354
696	NM_001732	BTN1A1	218	27	244
946	BC035359	SIGLEC6	96	28	123
1006	NM_006727	CDH10	239	22	260
1088	BC026263	CEACAM8	108	35	142
1088	BC026263	CEACAM8	286	35	320
1089	NM_001817	CEACAM4	120	35	154
1101	BC073974	CHAD	140	220	359
1117	BC011460	CHI3L2	364	27	390
1506	BC063475	CTRL	231	34	264
1519	BC049206	CTSO	213	108	320
1527	BC026183	CXorf2	410	1	410
1775	BC035205	DNASE1L2	122	21	142
1805	BC033736	DPT	184	18	201
1833	BC030958	DSPG3	304	19	322
1951	NM_001407	CELSR3	181	1764	1944
1952	NM_001408	CELSR2	309	1614	1922
1954	NM_001410	MEGF8	515	1	515
1954	NM_001410	MEGF8	143	804	946
2195	NM_005245	FAT	337	3788	4124
2195	NM_005245	FAT	253	570	822
2195	NM_005245	FAT	318	2809	3126
2196	NM_001447	FAT2	172	3773	3944
2267	BC007047	FGL1	227	79	305
2267	BC007047	FGL1	295	18	312
2563	BC033601	GABRD	232	17	248
2615	BC070079	LRRC32	201	200	400
2709	BC004379	GJB5	148	40	187
2765	BC074930	GML	141	18	158
2842	NM_006143	GPR19	67	1	67
2857	BC020678	GPR34	56	1	56
2859	NM_005301	GPR35	68	242	309
2882	BC032788	GPX7	169	19	187
3671	BC022478	ISLR	391	20	410
3742	BC069355	KCNA6	69	194	262
3755	BC046629	KCNG1	294	1	294
3770	BC035918	KCNJ14	137	300	436
4037	BC007408	LRP3	50	366	415
4045	BC033603	LSAMP	312	1	312
4238	BC026244	MFAP3	125	23	147
4239	BC062415	MFAP4	236	20	255
4311	NM_000902	MME	406	78	483
4685	BC052945	NCAM2	381	208	588
4832	BC000250	NME3	151	19	169
4885	BC034781	NPTX2	416	16	431
5098	BC019299	PCDHGC3	662	31	692
5099	NM_002589	PCDH7	270	30	299
5157	BC010927	PDGFRL	71	20	90
5276	BC027889	SERPINI2	378	28	405
5407	BC025784	PNLIPRP1	112	356	467
5407	BC025784	PNLIPRP1	450	18	467
5408	BC005969	PNLIPRP2	126	18	143
5408	BC005969	PNLIPRP2	452	18	469
5638	BC060833	PRRG1	63	21	83
5639	BC026032	PRRG2	86	1	86
5680	BC020711	PSG11	193	27	219
6039	BC020648	RNASE6	127	24	150
6398	BC017716	SECTM1	248	1	248
6425	BC050435	SFRP5	71	210	280
6511	BC028721	SLC1A6	109	154	262
6512	BC017242	SLC1A7	102	115	216
6610	BC000036	SMPD2	111	1	111
6676	NM_003116	SPAG4	248	190	437
6694	NM_006944	SPP2	183	29	211
6725	NM_080823	SRMS	93	120	212

EntrezGeneID	Accession Transcript	Symbol	Length_AA	Nterm_AA	Cterm_AA
6725	NM_080823	SRMS	259	230	488
7093	NM_012465	TLL2	126	429	554
7105	BC012389	TSPAN6	96	115	210
7180	NM_003296	CRISP2	223	21	243
7455	NM_173059	ZAN	470	1	470
7542	BC012814	ZFPL1	266	1	266
7732	BC053989	ZNF179	546	1	546
7844	BC035053	RNF103	298	28	325
7920	BC031839	BAT5	441	118	558
7922	BC000645	SLC39A7	104	30	133
7993	BC020694	UBXD6	214	57	270
8001	BC036086	GLRA3	219	34	252
8076	BC005901	MFAP5	146	28	173
8228	BC020746	PNPLA4	153	24	176
8228	BC020746	PNPLA4	229	25	253
8581	BC031330	LY6D	74	21	94
8749	NM_014237	ADAM18	366	16	381
8987	BC022301	GENX-3414	334	25	358
9019	BC007881	MPZL1	126	43	168
9027	BC012626	NAT8	162	66	227
9213	NM_004736	XPR1	105	369	473
9340	NM_004246	GLP2R	179	1	179
9350	BC069491	CER1	248	21	268
9358	BC036788	ITGBL1	93	23	115
9389	BC075071	SLC22A14	95	91	185
9399	BC034379	STOML1	398	1	398
9543	BC042054	PUNC	197	42	238
9603	BC068455	NFE2L3	120	295	414
9708	NM_014004	PCDHGA8	104	30	133
9708	NM_014004	PCDHGA8	105	243	347
9719	BC050544	ADAMTSL2	930	22	951
9723	NM_012431	SEMA3E	276	500	775
9766	BC064697	KIAA0247	81	40	120
9780	NM_014745	FAM38A	120	217	336
9813	BC002525	KIAA0494	403	93	495
9854	BC022219	TMEM24	124	37	160
9856	NM_014809	KIAA0319	250	19	268
9860	NM_014813	LRIG2	105	41	145
9860	NM_014813	LRIG2	100	498	597
9884	NM_014834	LRRC37A	302	36	339
9911	NM_014858	TMCC2	251	100	350
10003	NM_005467	NAALAD2	477	24	500
10004	NM_005468	NAALADL1	317	263	579
10082	NM_005708	GPC6	328	23	350
10162	BC065194	OACT5	69	302	370
10205	BC017774	EVA1	128	27	154
10218	BC001881	ANGPTL7	320	27	346
10218	BC001881	ANGPTL7	126	27	152
10246	NM_005835	SLC17A2	52	334	385
10330	BC065015	TMEM4	162	21	182
10343	NM_006071	PKDREJ	237	1730	1966
10410	BC006794	IFITM3	57	1	57
10446	BC034047	LRRN5	512	19	530
10462	BC039011	CLEC10A	316	1	316
10491	BC008745	CRTAP	305	26	330
10509	NM_198925	SEMA4B	674	43	716
10637	BC027663	LEFTY1	346	21	366
10648	BC062693	SCGB1D1	69	22	90
10695	NM_183010	TNRC5	278	1	278
10695	BC008698	TNRC5	248	31	278
10712	BC006493	C1orf2	475	194	668
10716	NM_001010848	NRG3	313	384	696
10748	BC069734	KLRA1	215	1	215
10752	NM_006614	CHL1	577	24	600
10867	BC071881	TSPAN9	97	107	203
10871	BC022279	CD300C	205	20	224
10875	BC033820	FGL2	227	209	435
10876	BC069407	FAM12A	122	26	147
10877	BC074957	CFHR4	63	23	85

EntrezGeneID	Accession Transcript	Symbol	Length_AA	Nterm_AA	Cterm_AA
10878	BC058009	CFHR3	312	19	330
10924	BC018999	SMPDL3A	431	23	453
10994	BC011722	ILVBL	604	29	632
10994	BC011722	ILVBL	133	500	632
11055	BC005223	ZPBP	322	30	351
11070	BC011948	TMEM115	107	245	351
11085	BC028372	ADAM30	662	27	688
11098	BC001278	PRSS23	360	24	383
11118	NM_007047	BTN3A2	271	1	271
11119	NM_194441	BTN3A1	271	1	271
11148	BC035971	HHLA2	327	20	346
11172	NM_007179	INSL6	193	21	213
11172	NM_007179	INSL6	114	55	168
11173	BC061631	ADAMTS7	510	452	961
11174	NM_197941	ADAMTS6	233	24	256
11174	NM_197941	ADAMTS6	259	601	859
11247	BC053561	NXPH4	285	24	308
11249	NM_007226	NXPH2	243	22	264
22862	NM_014923	FNDC3A	770	202	971
22955	BC009752	SCMH1	318	343	660
22990	NM_014982	PCNX	427	74	500
23023	BC039659	TMCC1	125	1	125
23052	BC026191	ENDOD1	321	22	342
23105	BC024300	FSTL4	584	22	605
23127	BC035672	GLT25D2	126	28	153
23251	NM_015206	KIAA1024	403	99	501
23302	BC009975	KIAA0523	80	145	224
23324	BC033307	MAN2B2	325	28	352
23333	XM_374422	-	131	447	577
23341	BC047363	DNAJC16	201	300	500
23341	BC047363	DNAJC16	762	1	762
23344	BC004998	FAM62A	68	1	68
23400	BC030267	ATP13A2	354	67	420
23415	NM_012285	KCNH4	227	1	227
23460	BC070125	ABCA6	185	1	185
23460	BC070125	ABCA6	136	50	185
23507	BC030607	LRRC8B	104	408	511
23544	NM_021115	SEZ6L	111	452	562
23554	BC031265	TSPAN12	114	111	224
23584	BC007313	VSIG2	220	23	242
23630	BC035330	KCNE1L	61	1	61
23659	BC062605	LYPLA3	360	33	412
23670	NM_013390	TMEM2	357	105	462
23731	NM_032012	C9orf5	99	256	354
25777	NM_015374	UNC84B	482	236	717
25777	NM_015374	UNC84B	365	236	600
25789	BC010445	C19orf4	244	25	268
25807	BC002705	RHBDD3	145	242	386
25849	BC013294	DKFZP564O0823	238	20	257
25890	BC030221	ABI3BP	391	29	419
25934	BC005935	NIPSNAP3A	229	19	247
25975	BC038567	EGFL6	179	80	258
25979	BC004126	DHRS7B	282	44	325
25987	BC020975	LRRC54	337	17	353
26020	NM_014045	LRP10	109	28	136
26033	NM_207303	ATRNL1	71	942	1012
26033	NM_207303	ATRNL1	81	690	770
26045	NM_015564	LRRTM2	390	33	422
26090	BC014049	ABHD12	76	1	76
26167	BC001186	PCDHB5	659	31	689
26253	BC000715	CLEC4E	219	1	219
26262	BC067105	TSPAN17	117	118	234
26608	BC012938	TBL2	418	29	446
27039	BC044581	PKD2L2	224	53	276
27055	BC001745	D4S234E	82	104	185
27163	BC006368	ASAHL	172	28	199
27290	NM_014471	SPINK4	51	30	80
27293	BC014444	SMPDL3B	261	21	281
27326	NM_032967	PCDH11X	328	23	350

EntrezGeneID	Accession Transcript	Symbol	Length_AA	Nterm_AA	Cterm_AA
28513	NM_021153	CDH19	230	21	250
28968	BC034948	SLC6A16	92	404	495
29090	BC000892	C18orf55	119	130	248
29801	BC053544	ZDHHC8	279	42	320
29920	BC014868	PYCR2	306	16	321
29953	NM_013381	TRHDE	194	62	255
29992	BC017812	PILRA	211	16	226
30010	BC047505	NXPH1	251	21	271
50487	BC025316	PLA2G3	491	19	509
50863	BC050716	HNT	284	33	316
51027	BC063405	BOLA1	117	21	137
51030	BC008430	FAM18B	53	73	125
51032	BC069455	ELA2B	254	16	269
51050	BC074932	PI15	82	19	100
51053	BC000039	FAM26B	62	124	185
51075	BC000665	TXNDC14	110	187	296
51097	BC026185	SCCPDH	429	1	429
51161	BC034766	C3orf18	59	1	59
51232	NM_016441	CRIM1	567	34	600
51241	BC001702	C14orf112	106	1	106
51286	BC034732	BM88	149	1	149
51309	BC002691	ARMCX1	453	1	453
51310	BC020565	SLC22A17	99	1	99
51337	BC001311	C8orf55	191	18	208
51348	NM_016523	KLRF1	169	63	231
51358	BC008742	TEX264	281	33	313
51385	NM_016089	ZNF589	373	1	373
51430	NM_014283	C1orf9	298	25	322
51567	NM_000575	TTRAP	271	1	271
51635	BC000637	DHRS7	311	29	339
51661	BC009711	FKBP7	199	24	222
51669	BC015012	TMEM66	55	195	259
51705	BC017781	EMCN	161	18	178
51768	BC005176	TM7SF3	276	21	296
51816	BC051755	CECR1	483	29	511
53822	BC018619	FXYD7	80	1	80
53942	NM_014361	CNTN5	254	18	271
54097	BC057629	FAM3B	206	30	235
54360	BC031391	CYTL1	114	23	136
54470	BC007677	ARMCX6	280	21	300
54471	BC008327	RP5-1104E15.5	463	1	463
54504	BC016836	CPVL	447	22	468
54510	NM_019035	PCDH18	671	27	697
54587	BC006213	MXRA8	322	19	340
54662	BC032998	MANSC1	131	255	385
54716	NM_020208	SLC6A20	94	299	392
54757	NM_017565	FAM20A	301	100	400
54762	BC028972	GRAMD1C	213	1	213
54869	BC004907	EPS8L1	536	16	551
54894	NM_017763	RNF43	301	400	700
54929	BC005210	FLJ20422	53	387	449
54947	BC002472	AYTL1	544	1	544
54964	BC002469	C1orf56	319	23	341
54968	BC002748	TMEM70	260	1	260
54991	NM_017891	C1orf159	211	44	254
54996	BC011973	MOSC2	296	40	335
55013	BC002633	FLJ20647	248	1	248
55026	NM_017938		110	111	220
55028	BC005005	C17orf80	509	1	509
55092	BC000593	TMEM51	253	1	253
55104	NM_018038		87	30	116
55113	BC028564	XKR8	93	68	160
55129	BC038655	TMEM16K	236	1	236
55146	BC001239	ZDHHC4	72	121	192
55151	BC000049	TMEM38B	61	231	291
55177	BC063844	FAM82C	192	53	244
55177	BC063844	FAM82C	439	32	470
55194	BC005241	C1orf78	165	1	165
55218	BC001962	EXDL2	477	20	496

EntrezGeneID	Accession Transcript	Symbol	Length_AA	Nterm_AA	Cterm_AA
55216	BC001962	EXDL2	363	134	496
55246	BC006320	C1orf75	350	1	350
55253	NM_018264	RSAFD1	179	232	410
55253	NM_018264	RSAFD1	141	592	732
55260	BC016919	TMEM143	234	45	278
55268	BC044574	ECHDC2	121	105	225
55273	BC010128	TMEM100	134	1	134
55281	BC020942	TMEM140	45	37	81
55366	NM_018490	LGR4	277	24	300
55471	BC004548	PRO1853	423	19	441
55471	BC004548	PRO1853	423	19	441
55627	NM_017751	FLJ20297	70	729	798
55711	BC022267	MLSTD1	71	285	355
55739	NM_018210	FLJ10769	171	30	200
55744	BC056884	FLJ10803	115	32	146
55836	BC029130	C6orf35	141	1	141
55847	BC008474	C10orf70	108	1	108
55848	BC008212	C9orf46	75	73	147
55850	BC008455	MDS032	232	1	232
55852	BC036672	TEX2	315	518	832
55852	BC036672	TEX2	142	1	142
55863	BC038933	TMEM126B	200	1	200
55879	NM_018558	GABRQ	248	21	268
55997	BC074826	CFC1	199	25	223
56005	BC010129	C19orf10	142	32	173
56053	BC038842	C1orf91	164	1	164
56097	NM_018929	PCDHGC5	104	351	454
56096	NM_018928	PCDHGC4	109	134	242
56096	NM_018928	PCDHGC4	110	456	565
56096	NM_018928	PCDHGC4	663	30	692
56100	NM_018926	PCDHGB6	301	300	600
56102	NM_018924	PCDHGB3	103	31	133
56103	NM_018923	PCDHGB2	105	243	347
56103	NM_018923	PCDHGB2	105	348	452
56103	NM_018923	PCDHGB2	661	31	691
56104	NM_018922	PCDHGB1	106	566	671
56104	NM_018922	PCDHGB1	109	131	239
56104	NM_018922	PCDHGB1	659	29	687
56105	NM_032091	PCDHGA11	301	300	600
56106	NM_032090	PCDHGA10	201	300	500
56107	NM_018921	PCDHGA9	241	60	300
56108	NM_018920	PCDHGA7	301	200	500
56109	NM_018919	PCDHGA6	104	30	133
56109	NM_018919	PCDHGA6	663	30	692
56109	NM_018919	PCDHGA6	109	134	242
56111	NM_032053	PCDHGA4	105	29	133
56111	NM_032053	PCDHGA4	105	243	347
56112	NM_018916	PCDHGA3	104	30	133
56113	NM_018915	PCDHGA2	105	29	133
56113	NM_018915	PCDHGA2	109	134	242
56113	NM_018915	PCDHGA2	664	29	692
56114	NM_018912	PCDHGA1	105	29	133
56114	NM_018912	PCDHGA1	109	134	242
56122	NM_018934	PCDHB14	659	30	688
56125	NM_018931	PCDHB11	663	26	688
56241	BC033107	SUSD2	167	447	613
56241	BC033107	SUSD2	149	285	433
56244	NM_019602	BTNL2	301	43	343
56246	BC062721	MRAP	114	59	172
56255	BC063430	CPXM	252	20	271
56674	BC040124	TMEM98	165	34	198
56884	BC036502	FSTL5	241	23	263
56914	NM_020157	OTOR	109	20	128
56920	NM_020163	SEMA3G	445	58	503
56926	BC013263	NCLN	166	355	520
56928	BC028391	SPPL2B	148	25	172
56934	BC020577	CA10	308	21	328
56935	BC031564	C11orf75	59	1	59
56951	BC020675	C5orf15	150	50	199

EntrezGeneID	Accession Transcript	Symbol	Length_AA	Nterm_AA	Cterm_AA
56967	NM_020215	C14orf132	173	1	173
56967	NM_020215	C14orf132	131	1	131
56975	BC040074	FAM20C	216	355	570
56983	BC048810	C3orf9	133	260	392
57003	BC008905	CCDC47	464	20	483
57003	BC008905	CCDC47	346	130	475
57003	BC008905	CCDC47	111	20	130
57094	BC033684	CPA5	280	145	424
57101	NM_020373	TMEM16B	359	1	359
57104	BC017280	PNPLA2	478	27	504
57150	BC070260	C6orf162	97	1	97
57151	BC054481	LYZL6	129	20	148
57153	BC040556	SLC44A2	174	57	230
57171	BC033686	DOLPP1	56	183	238
57181	NM_020342	SLC39A10	378	26	403
57181	NM_020342	SLC39A10	229	27	255
57186	NM_207517	ADAMTSL3	84	1302	1385
57191	NM_020633	VN1R1	51	1	51
57214	BC020256	KIAA1199	963	30	992
57406	BC074741	LRTM1	120	51	170
57453	NM_020693	DSCAML1	302	1350	1651
57484	BC101992	RNF150	175	140	314
57486	NM_020728	FAM62B	167	609	775
57512	NM_020752	GPR158	118	301	418
57544	NM_020784	KIAA1344	226	600	825
57552	BC028734	AADACL1	381	60	440
57574	NM_020814	MARCH4	218	18	235
57582	XM_029962	KCNT1	189	597	785
57582	XM_029962	KCNT1	213	999	1211
57586	NM_020826	SYT13	397	30	426
57611	NM_020851	ISLR2	133	241	373
57622	XM_290842	LRFN1	771	1	771
57622	XM_290842	LRFN1	212	1	212
57642	BC043163	COL20A1	173	186	358
57642	BC043163	COL20A1	203	835	1037
57653	NM_020893	KIAA1529	489	1158	1646
57670	XM_371956	KIAA1549	211	1	211
57670	XM_371956	KIAA1549	613	1264	1876
57715	NM_017893	SEMA4G	667	17	683
57719	NM_020959	TMEM16H	246	1	246
57720	NM_020960	GPR107	226	39	264
57722	NM_020962	NOPE	298	141	438
57756	NM_020974	SCUBE2	245	120	364
57756	NM_020974	SCUBE2	161	644	804
57826	NM_021185	C19orf15	251	250	500
57864	BC048285	TSCOT	77	1	77
58189	BC029159	WFDC1	190	31	220
58496	NM_021221	LY6G5B	138	64	201
58527	BC014953		81	1	81
58985	BC029273	IL22RA1	325	250	574
59084	BC027615	ENPP5	408	22	429
59284	BC069332	CACNG7	81	20	100
59307	BC025953	SIGIRR	270	141	410
60314	BC051871	C12orf10	103	20	122
60401	BC034919	EDA2R	139	1	139
60484	BC029864	HAPLN2	315	26	340
60492	BC014573	MDS025	230	1	230
60598	NM_022358	KCNK15	59	23	81
60686	BC014299	C14orf93	295	6	300
63027	NM_021945	C6orf85	62	84	145
63895	NM_022068	FAM38B	118	242	359
63895	NM_022068	FAM38B	92	360	451
64063	BC009726	PRSS22	286	32	317
64094	BC047583	SMOC2	426	21	446
64100	BC015598	ELSPBP1	200	24	223
64109	NM_001012268	CRLF2	120	1	120
64115	BC020568	C10orf54	160	34	193
64115	BC020568	C10orf54	280	32	311
64123	BC025721	ELTD1	346	1	346

EntrezGeneID	Accession Transcript	Symbol	Length_AA	Nterm_AA	Cterm_AA
64129	BC064633	TINAGL1	261	21	281
64150	OTTHUMT00000072656	DIO3OS	122	15	136
64175	NM_022356	LEPRE1	114	22	135
64222	BC011746	TOR3A	373	25	397
64285	BC014425	RHBDF1	552	1	552
64409	BC069645	WBSCR17	129	27	155
64420	BC060770	SUSD1	300	201	500
64420	BC060770	SUSD1	257	501	757
64430	NM_022495	C14orf135	131	413	543
64579	NM_022569	NDST4	225	36	260
64748	BC009378	LPPR2	94	35	128
64753	NM_022742	NAG6	230	1	230
64753	NM_022742	NAG6	342	650	991
64753	NM_022742	NAG6	90	345	434
64755	BC009308	C16orf58	247	1	247
64757	BC010619	MOSC1	301	37	337
64806	BC069565	IL17E	149	29	177
64838	BC032725	FNDC4	164	1	164
64840	BC019060	PORCN	113	220	332
64921	BC063284	CASD1	310	1	310
64922	NM_022901	LRRC19	246	25	270
65980	BC041590	BRD9	463	19	481
65983	BC008590	GRAMD3	342	1	342
65990	BC001181	C16orf24	193	43	235
65990	BC001181	C16orf24	193	43	235
65992	BC000643	C20orf116	97	218	314
66000	BC000568	TMEM108	438	30	467
66000	BC000568	TMEM108	111	51	161
66004	NM_177458	LYNX1	76	22	97
66005	BC000001	CHID1	372	22	393
78989	BC000078	COLEC11	246	26	271
78992	BC013014	YIPF2	124	1	124
78997	BC009014	GDAP1L1	339	1	339
79022	BC000854	TMEM106C	85	110	194
79037	BC001129	MGC2463	170	1	170
79041	BC001195	TMEM38A	67	233	299
79135	BC002333	MGC4825	198	1	198
79143	BC003164	LENG4	158	268	425
79153	BC002714	GDPD3	97	164	260
79154	NM_024308	MGC4172	260	1	260
79154	NM_024308	MGC4172	228	33	260
79157	BC002753	ET	51	336	386
79174	BC050675	CRELD2	330	24	353
79411	NM_024506	GLB1L	627	28	654
79412	BC003533	KREMEN2	395	26	420
79415	BC003595	C17orf62	150	36	187
79584	BC067526	FLJ12684	466	24	489
79600	BC040113	FLJ21127	552	22	573
79630	BC017761	C1orf54	115	17	131
79639	BC064520	TMEM53	277	1	277
79651	BC016034	RHBDF2	204	450	653
79669	BC017064	C3orf52	124	27	150
79701	BC023602	FLJ22222	331	1	331
79713	NM_024660	TMEM149	142	23	164
79714	BC011993	CCDC51	380	1	380
79742	NM_024689	CXorf36	152	31	182
79752	BC036067	C1orf115	142	1	142
79770	BC032568	C5orf14	292	32	323
79789	BC068482	CLMN	234	743	976
79815	NM_024759	NPAL2	86	1	86
79820	NM_024764	C14orf161	170	25	194
79827	BC009371	ASAM	216	18	233
79844	BC032000	ZDHHC11	86	91	176
79847	NM_024789	C10orf77	55	212	266
79852	NM_024794	ABHD9	230	124	353
79853	BC035754	TM4SF20	61	120	180
79867	BC009112	C12orf36	274	171	444
79875	NM_024817	THSD4	110	26	135
79876	BC009737	UBE1DC1	70	1	70

EntrezGeneID	Accession Transcript	Symbol	Length_AA	Nterm_AA	Cterm_AA
79879	BC017693	FLJ22349	207	23	229
79883	BC057786	FLJ23447	486	27	512
79887	BC063561	FLJ22662	520	33	552
79888	NM_024830	AYTL2	152	383	534
79905	BC036205	TMC7	168	1	168
79906	NM_024850	BTNL8	330	18	347
79956	NM_024896	KIAA1815	231	674	904
79956	NM_024896	KIAA1815	204	1	204
79974	NM_024913	FLJ21986	266	35	300
79974	NM_024913	FLJ21986	400	301	700
79974	NM_024913	FLJ21986	326	701	1026
80006	BC030803	FLJ23235	111	50	160
80020	BC027716	RP5-1119A7.4	322	31	352
80023	BC001963	C20orf96	204	1	204
80031	NM_153616	SEMA6D	569	20	588
80221	BC014123	FLJ20920	440	101	540
80341	BC034415	BPIL1	439	20	458
80346	BC013048	REEP4	195	63	257
80350	NM_145727	LPAL2	271	1	132
80381	BC062581	CD276	466	28	493
80736	BC014659	SLC44A4	169	59	227
80740	BC036302	LY6G6C	107	19	125
80761	BC004304	UPK3B	291	30	320
80762	BC004317	NDFIP1	116	1	116
80763	BC004336	C12orf39	90	27	116
80864	NM_030652	EGFL8	79	34	112
80864	NM_030652	EGFL8	150	144	293
81025	BC051675	GJA10	283	233	515
81037	BC025305	CRR9	151	33	183
81491	BC067469	GPR63	81	1	81
81533	BC006321	ITFG1	536	30	565
81542	BC036460	TXNDC	261	20	280
81562	BC067265	LMAN2L	304	45	348
81575	NM_030817	APOLD1	68	181	248
81579	BC017218	PLA2G12A	167	23	189
81671	BC009758	TMEM49	140	133	272
81792	BC058841	ADAMTS12	203	27	229
81832	BC050329	NETO1	175	22	196
81833	BC029488	SPACA1	188	29	216
83440	BC006112	ADPGK	476	22	497
83445	BC033854	GSG1	96	36	131
83539	BC025764	CHST9	409	30	438
83590	BC000936	C7orf21	165	31	195
83636	NM_031448	C19orf12	86	54	141
83643	BC051334	CCDC3	249	22	270
83690	BC020514	CRISPLD1	185	21	205
83716	BC063012	CRISPLD2	263	26	288
83716	BC063012	CRISPLD2	161	288	448
83729	BC005161	INHBE	332	19	350
83787	BC003586	SVH	281	28	308
83850	BC037292	FAM62C	429	73	501
83882	BC032602	TSPAN10	177	179	355
83886	BC034294	PRSS27	269	22	290
83888	BC025720	KSP37	205	19	223
83985	BC038961	SPIN1	96	1	96
83986	BC032112	ITFG3	169	72	240
83999	BC063767	KREMEN1	439	20	458
84063	BC064925	KIRREL2	357	20	376
84066	BC034972	C1orf49	215	1	215
84102	BC036734		75	1	75
84133	XM_290972	ZNRF3	137	700	836
84133	XM_290972	ZNRF3	211	250	460
84141	BC016157	FLJ13391	152	1	152
84179	BC030246	MFSD7	118	442	559
84188	BC017377	MLSTD2	465	1	465
84189	NM_032229	SLITRK6	585	26	610
84189	NM_032229	SLITRK6	136	475	610
84189	NM_032229	SLITRK6	140	224	363
84197	NM_032237	FLJ23356	350	1	350

EntrezGeneID	Accession Transcript	Symbol	Length_AA	Nterm_AA	Cterm_AA
84197	NM_032237	FLJ23356	307	44	350
84216	BC060798	TMEM117	98	417	514
84233	BC007875	TMEM126A	195	1	195
84236	NM_032276	RHBDD1	55	126	180
84239	NM_032279	ATP13A4	168	57	224
84258	BC031067	SYT3	517	74	590
84273	BC004894	C4orf14	296	24	319
84277	BC005056	WBSCR18	186	22	207
84279	BC005069	C2orf7	167	22	188
84287	BC008074	ZDHHC16	240	138	377
84293	BC005871	C10orf58	229	1	229
84314	BC070231	TMEM107	59	1	59
84329	BC007277	HVCN1	65	1	65
84334	BC007412	C14orf153	176	18	193
84417	BC021742	ECRG4	116	33	148
84439	NM_032425	KIAA1622	301	100	400
84466	NM_032446	MEGF10	96	315	410
84514	BC022784	LGP1	513	18	530
84519	BC033010	ACRBP	519	25	543
84623	NM_032531	KIRREL3	515	21	535
84631	NM_032539	SLITRK2	601	21	621
84631	NM_032539	SLITRK2	117	87	203
84631	NM_032539	SLITRK2	137	203	339
84659	BC074960	RNASE7	132	25	156
84681	BC047737	HINT2	145	19	163
84696	BC039575	ABHD1	189	217	405
84709	NM_032623	OSAP	268	1	268
84709	NM_032623	OSAP	205	64	268
84804	BC006242	MGC11332	69	219	287
84833	BC007087	USMG5	56	1	56
84866	BC051841	TMEM25	297	26	322
84868	BC063431	HAVCR2	181	21	201
84870	BC022367	RSPO3	252	21	272
84886	BC025740	SPPL2A	145	25	169
84894	BC011057	LRRN6A	519	41	559
84896	BC012685	PLXDC2	375	30	404
84910	NM_032824	TMEM87B	173	43	215
84910	NM_032824	TMEM87B	106	43	148
84918	BC043141	LRP11	413	36	450
84957	BC017279	TNFRSF19L	137	26	162
84976	NM_032890	DISP1	288	211	498
84978	BC053647	FRMD5	48	523	570
85315	BC030664	PAQR8	55	264	318
85318	NM_182481	BAGE3	93	17	109
85319	NM_182482	BAGE2	93	17	109
85413	BC047565	SLC22A16	110	44	153
85450	BC070108	KIAA1754	151	350	500
85455	NM_033510	DISP2	234	725	958
89782	NM_033029	LMLN	531	1	531
89872	BC074897	AQP10	94	208	301
89932	BC042057	PAPLN	112	18	129
90141	NM_145231	C14orf143	163	1	163
90141	NM_145231	C14orf143	138	22	159
90199	NM_130896	WFDC8	204	38	241
90231	BC035033	KIAA2013	222	41	262
90273	BC012001	CEACAM21	158	34	191
90313	BC001593	TP53I13	284	1	284
90342	XM_031009		124	582	705
90342	XM_031009		84	233	316
90342	XM_031009		706	1	706
90407	BC019684	TMEM41A	51	18	68
90527	BC029619	NIP	109	75	183
90693	BC012427	LOC90693	105	36	140
90871	BC009510	C9orf123	116	1	116
90952	BC016868	ESAM	219	29	247
90990	BC017311	KIFC2	326	69	394
91147	BC054338	TMEM67	201	300	500
91161	NM_207308	NUP210L	93	790	882
91181	NM_207308	NUP210L	210	1	210

EntrezGeneID	Accession Transcript	Symbol	Length_AA	Nterm_AA	Cterm_AA
91252	BC019016	SLC39A13	112	39	150
91304	BC008957	C19orf6	66	1	66
91452	BC030555	ACBD5	459	1	459
91584	BC028744	PLXNA4B	501	22	522
91689	BC024237	LOC91689	107	1	107
91775	BC009431	FAM55C	232	33	264
91851	BC002909	CHRDL1	430	27	456
91862	BC011893	MARVELD3	270	1	270
91937	BC008968	TIMD4	187	128	314
92126	NM_032160	C18orf4	115	31	145
92126	NM_032160	C18orf4	152	593	744
92255	NM_001007527	LMBRD2	145	551	695
92270	BC093982	LOC92270	224	1	224
92270	BC093982	LOC92270	57	168	224
92305	NM_138365	TMEM129	118	115	232
92370	BC035834	ACPL2	455	26	480
92691	BC008604	LOC92691	297	1	297
92691	BC008604	LOC92691	159	1	159
92737	BC035009	DNER	320	34	353
92747	BC008429	C20orf114	455	30	484
92949	BC030262	ADAMTSL1	411	29	439
93109	NM_001011655	TMEM44	62	1	62
93109	NM_001011655	TMEM44	67	112	178
93129	BC006126	MGC13024	68	177	244
93210	BC010652	PERLD1	80	20	99
93978	NM_001007033	CLEC6A	166	44	209
94031	BC034390	HTRA3	437	17	453
94101	BC005200	ORMDL1	57	44	100
94120	BC065612		125	1	125
94240	BC023660	EPSTI1	410	1	410
112609	BC039855	C6orf117	205	1	205
112770	BC018757	C1orf85	336	36	371
112817	NM_138413	C10orf65	327	1	327
113235	BC010691	HCP1	84	1	84
113277	BC012139	TMEM106A	146	117	262
113277	BC012139	TMEM106A	146	117	262
113791	BC011049	MGC17330	147	22	168
114659	NM_052868	LRRC37B	210	1	210
114780	NM_052892	PKD1L2	205	1141	1345
114783	XM_055866	LMTK3	295	865	1160
114794	NM_052906	KIAA1904	250	23	272
114796	BC051738	SLITRK1	114	262	375
114897	BC021553	C1QTNF1	256	26	281
114904	BC020551	C1QTNF6	232	47	278
114908	BC032296	TMEM123	142	27	168
114915	NM_053000	TIGA1	120	1	120
114926	BC013035	C8orf40	100	1	100
114990	BC068575	VASN	554	23	576
115019	NM_052934	SLC26A9	296	496	791
115111	NM_052832	SLC26A7	194	463	656
115123	BC047569	MARCH3	253	1	253
115330	BC014241	GPR146	123	211	333
115416	BC012331	C7orf30	105	91	195
116150	BC066910	C6orf68	102	24	125
116211	BC013113	TM4SF19	54	120	173
116236	BC012476	LOC116236	446	23	468
116254	BC014320	C6orf72	233	29	261
116969	BC014577	ART5	271	22	292
117144	BC032950	CATSPER1	442	1	442
118813	BC030621	ZFYVE27	204	208	411
118881	BC023663	COMTD1	230	33	262
118881	BC023663	COMTD1	126	137	262
118932	BC021671	ANKRD22	191	1	191
118987	BC028375	PDZD6	422	25	446
119395	BC043367	FAM26A	152	199	350
119467	BC029478	TMEM12	58	36	93
119587	BC036789	CPXM2	261	30	290
120224	BC016153	TMEM45B	70	23	92
120939	NM_153022	C12orf59	121	43	163

EntrezGeneID	Accession Transcript	Symbol	Length_AA	Nterm_AA	Cterm_AA
121227	NM_153377	LRIG3	168	24	191
121256	NM_133448	TMEM132D	475	30	504
121506	BC030218	C12orf46	151	123	273
121601	NM_178826	TMEM16D	166	755	920
121665	BC073910	UNQ1887	50	211	260
121793	BC029689	MGC35169	151	1	151
122258	BC016750	LOC122258	171	25	195
122618	BC015003	PLD4	147	45	191
122651	BC025410	RNASE11	183	17	199
122651	BC025410	RNASE11	114	17	130
123041	BC069653	SLC24A4	179	229	407
124220	BC009722	LOC124220	156	17	172
124446	BC050051	LOC124446	162	43	204
124446	BC050051	LOC124446	108	43	150
124565	BC014642	MGC15523	383	398	780
124599	BC028091	CD300LB	238	1	238
124626	BC043152	ZPBP2	300	17	316
124936	BC051697	CYB5D2	242	23	264
124944	BC040036	C17orf49	172	1	172
124944	BC040036	C17orf49	104	1	104
125170	BC014973	SMCR7	407	48	454
125206	BC039868	SLC5A10	165	131	295
125206	BC039868	SLC5A10	57	239	295
125228	BC022410	C16orf19	132	1	132
125931	NM_196444	CEACAM20	249	200	448
125968	BC009557	P117	101	18	118
126259	BC015655	MGC23244	260	23	282
126364	BC071640	LRRC25	145	21	165
126526	BC027935	FLJ36888	355	1	355
126969	BC053877	SLC44A3	165	1	165
127579	NM_144622	DCST2	69	253	321
127700	BC018069	C1orf102	353	27	379
128240	BC056917	APOA1BP	237	24	260
128414	BC041812	C20orf56	208	1	208
128434	BC033818	C20orf102	180	25	204
128497	BC039607	C20orf165	227	1	227
128497	BC039607	C20orf165	151	1	151
128646	BC033602	SIRPD	105	31	135
128822	NM_001008693	CST9	130	30	159
128861	BC066354	C20orf71	198	21	218
129060	BC046358	EMID1	153	29	181
129530	BC029126	LOC129530	175	20	194
129804	NM_153214	FLJ37440	129	140	268
129804	NM_153214	FLJ37440	184	136	319
130814	BC027625	PQLC3	57	115	171
130814	BC027625	PQLC3	183	20	202
130827	NM_144632	FLJ30294	73	41	113
131096	NM_144633	KCNH6	223	1	223
131177	BC015369	FAM3D	197	28	224
131375	BC016747	LYZL4	128	19	146
131566	BC029658	DCBLD2	462	21	482
131578	NM_130830	LRRC15	153	21	173
131873	XM_067585	LOC131873	199	1103	1301
132228	BC028000	-	164	1	164
132720	BC022534	FLJ39370	105	1	105
132724	NM_182502	TMPRSS11B	226	185	410
133308	BC047447	LOC133308	49	257	305
133308	BC047447	LOC133308	115	375	489
133418	BC059398	EMB	232	32	263
133482	BC034976	SLCO6A1	65	541	605
133688	BC068446	UGT3A1	201	250	450
135886	BC030643	WBSCR28	265	1	265
135927	BC014596	C7orf34	94	29	122
136242	NM_001008270	LOC136242	216	20	235
136242	NM_001008270	LOC136242	132	40	171
136263	BC017587	LOC136263	244	1	244
136306	BC036795	LOC136306	52	253	304
136541	NM_001001317	TRY1	184	41	224
137797	NM_205545	LYPD2	103	23	125

EntrezGeneID	Accession Transcript	Symbol	Length_AA	Nterm_AA	Cterm_AA
137797	NM_205545	LYPD2	103	23	125
137835	BC062592	TMEM71	276	1	276
138065	BC013036	RNF183	192	1	192
138311	BC032097	FAM69B	251	150	400
140456	BC069340	ASB11	97	227	323
140683	BC065726	C20orf70	231	19	249
140832	NM_080753	WFDC10A	59	21	79
140870	NM_080827	WFDC6	62	25	86
140902	NM_176491	R3HDML	140	67	206
142683	NM_080878	ITLN2	290	36	325
143162	NM_152428	FRMPD2	235	314	548
143162	NM_152428	FRMPD2	390	750	1139
143162	NM_152428	FRMPD2	298	18	315
143282	BC025966	C10orf13	99	27	125
143662	BC058007	MUC15	213	24	236
143903	BC025407	LAYN	204	24	227
144195	BC060766	SLC2A14	67	204	270
144321	BC029557	MGC39497	211	43	253
144383	XM_084845		57	1	57
145264	BC040857	SERPINA12	395	19	414
145407	NM_001001672	C14orf37	222	26	249
145748	BC064545	LYSMD4	98	120	217
145748	BC064545	LYSMD4	119	1	119
145864	BC062320	HAPLN3	343	18	360
145942	BC029221	TMCO5	288	1	288
145957	BC017568	NRG4	60	1	60
146378	XM_375333		100	1	100
146395	BC016460	UNQ5831	176	1	176
146429	XM_370997	LOC146429	100	49	148
146433	BC029804	MGC34647	222	21	242
146556	NM_152459	MGC45438	201	100	300
146802	BC050578	FLJ31196	86	490	577
146852	NM_153007	ODF4	77	1	77
146894	BC025395	CD300LG	233	19	251
147007	BC033113	C17orf32	208	1	208
147015	NM_144683	MGC23280	327	1	327
147138	NM_152468	TMC6	60	139	198
147172	NM_207323	DKFZp667M2411	87	1	87
147381	BC035789	CBLN2	175	50	224
147495	BC053324	APCDD1	467	26	492
147645	XM_085831	LOC147645	94	302	395
147665	BC033933	C19orf18	192	24	215
147719	NM_173506	LYPD4	220	27	246
147719	NM_173506	LYPD4	110	26	135
147744	NM_139172	MDAC1	177	1	177
147798	BC025323	TMC4	94	443	536
147920	NM_001002915	IGFL2	95	29	123
147991	BC029162	DPY19L3	102	1	102
148206	XM_496299	ZNF714	205	23	227
148753	NM_173509	C1orf76	141	27	167
148808	BC036549	MFSD4	63	242	304
148811	BC063477	FLJ32569	477	26	502
149233	NM_144701	IL23R	318	36	353
149421	NM_152497		188	1	188
149466	BC041633	C1orf210	113	1	113
149830	NM_177549	PRNT	77	18	94
150165	BC045548	XKR3	51	115	165
150372	BC038241	NFAM1	111	1	111
150696	NM_144707	PROM2	248	177	424
150771	BC034503	KIAA1754L	428	136	563
151056	BC042674	PLS1	379	1	379
151393	BC024243	FAM62A	111	300	410
151473	BC065524	SLC15A14	167	148	314
151647	BC031566	FAM19A4	105	36	140
152002	BC039067	C3orf21	121	30	150
152028	BC063696	FNDC6	102	36	137
152404	BC034411	IGSF11	226	19	244
152519	BC067661	NPAL1	55	1	55
152519	BC067661	NPAL1	53	358	410

EntrezGeneID	Accession Transcript	Symbol	Length_AA	Nterm_AA	Cterm_AA
152816	NM_178457	FLJ23657	107	24	130
154141	BC045695	OACT1	238	134	371
154197	NM_173516	PNLDC1	360	5	364
154197	NM_173516	PNLDC1	157	364	520
154467	XM_166346	C5orf129	97	1	97
157708	XM_088367	SEC11L2	136	162	297
157869	BC042877	RPESP	111	15	125
158062	BC062746	LCN6	143	21	163
158521	BC034320	FMR1NB	96	91	186
158584	BC073922	RP11-479E16.1	351	100	450
158763	NM_144967	RP13-102H20.1	357	41	397
158763	NM_144967	RP13-102H20.1	151	397	547
159989	BC050325	CCDC67	125	1	125
159989	BC050325	CCDC67	356	1	356
160065	BC039863	PATE	106	21	126
160335	NM_152568	TMTC2	64	329	392
160418	NM_181763	TMTC3	111	804	914
160697	BC052243	GPR160	151	22	172
161176	BC061899	C14orf49	66	1	66
161198	BC031567	CLEC14A	153	201	353
161198	BC031567	CLEC14A	201	1	201
162387	BC040487	FLJ35773	63	456	518
162514	NM_145068	TRPV3	250	1	250
162540	BC025401	IMP5	157	28	184
163486	NM_144977	DENND1B	177	95	271
163486	NM_144977	DENND1B	302	95	396
163882	NM_152609	C1orf71	415	1	415
163882	NM_152609	C1orf71	245	415	659
164091	BC034015	PAQR7	55	260	314
164153	BC058929	USL4B	155	20	174
164312	BC027720	C20orf75	156	365	520
164656	BC039062	TMPRSS5	386	76	461
165215	BC060872	KIAA1946	354	1	354
165631	NM_152615	PARP15	425	19	444
167681	BC037170	PRSS35	271	133	403
167681	BC037170	PRSS35	393	21	413
168433	BC022038	RNF133	155	34	188
168507	NM_138295	PKD1L1	212	2306	2517
168667	BC060868	BMPER	319	39	357
169611	NM_182487	OLFML2A	256	82	337
169693	BC029760	C9orf71	146	25	170
170392	NM_152635	OIT3	200	22	221
170487	OTTHUMT00000076713	C20orf134	368	29	396
170575	BC040736	GIMAP1	272	1	272
192111	BC008196	PGAM5	226	30	255
192286	BC000567	HIGD2A	106	1	106
195814	BC064525	RDHE2	285	25	309
196264	NM_198275	LOC196264	127	32	158
196463	BC030618	LOC196463	190	400	589
196740	BC041414	C10orf72	298	23	320
196740	BC041414	C10orf72	156	23	178
196792	NM_152644	FAM24B	65	30	94
196996	XM_113796	GRAMD2	139	1	139
197322	NM_174917	LOC197322	184	17	200
197322	NM_174917	LOC197322	176	401	576
199675	NM_174918	MCEMP1	187	1	187
199731	NM_145296	IGSF4C	300	24	323
199953	NM_001010866	RP13-15M17.2	88	237	324
200232	XM_371401	C20orf106	214	1	214
200383	BC015442	-	126	426	551
200504	NM_182536	GDDR	164	21	184
200634	NM_173853	KRTCAP3	126	115	240
200634	NM_173853	KRTCAP3	204	37	240
201156	BC011952	FAM18B2	276	1	276
201164	BC031263	LOC201164	230	23	252
201164	BC018995	FAM70B	85	110	194
201229	XM_375430	LOC201229	78	1	78
201243	BC031286	C17orf74	360	49	408
201243	BC031286	C17orf74	109	300	408

EntrezGeneID	Accession Transcript	Symbol	Length_AA	Nterm_AA	Cterm_AA
201305	BC023646	MGC29671	115	271	385
201633	NM_173799	VSIG9	229	16	244
201780	BC019065	SLC10A4	115	323	437
202915	BC026694	MGC9712	126	1	126
203062	NM_145003	TSNARE1	53	421	483
203062	NM_145003	TSNARE1	110	1	110
203069	XM_114618		101	50	150
203074	NM_196464	UNQ9391	334	19	352
203074	NM_196464	UNQ9391	132	19	150
203260	BC018758	MGC31967	259	25	283
203260	BC018758	MGC31967	201	83	283
203328	BC014601	SUSD3	106	1	106
203413	BC062223	RP3-452H17.2	113	1	113
203562	BC029675	TMEM31	166	1	166
204962	BC028743	SLC44A5	179	62	240
205327	BC036456	FLJ38973	245	25	269
206938	NM_152702	C9orf94	242	21	262
206938	NM_152702	C9orf94	91	99	189
207063	BC032340	DHRSX	300	31	330
219348	XM_166090	PLAC9	97	1	97
219699	NM_170744	UNC5B	350	26	375
219833	BC025756	C11orf45	123	23	145
219928	BC016964	MRGPRF	80	264	343
219931	BC063008	TPCN2	124	312	435
219938	BC058039	SPATA19	151	17	167
219990	BC036256	TMEM122	141	18	158
220323	BC047726	OAF	246	28	273
220686	BC020225	LOC220686	319	31	349
221035	BC068557	REEP3	174	82	255
221091	BC053902	LOC221091	216	23	238
221188	BC032401	GPR114	248	1	248
221191	BC057843	Klkbl4	366	18	383
221395	BC066121	GPR116	260	22	281
221786	NM_145111	C7orf38	573	1	573
221786	NM_145111	C7orf38	280	15	294
221955	BC027603	LOC221955	518	155	672
222008	NM_182546	MGC33530	220	24	243
222008	NM_182546	MGC33530	101	143	243
222223	NM_152748	KIAA1324L	228	374	601
222537	NM_153612	HS3ST5	315	32	346
222663	BC052263	SCUBE3	377	20	396
222865	BC030793	TMEM130	163	177	339
245812	BC032339	MGC40499	216	33	248
245911	NM_001002035	DEFB108B	52	22	73
252839	BC001106	TMEM9	71	113	183
253012	NM_196151	LOC253012	89	137	225
254228	BC032556	C6orf188	191	119	309
254359	BC057633	ZDHHC24	68	217	284
254531	NM_153613	AGPAT7	306	1	306
254531	NM_153613	AGPAT7	244	63	306
254773	NM_175735	LYG2	193	20	212
255022	BC036193	FAM26C	101	220	320
255057	BC028156	C19orf26	386	62	447
255104	BC053600	-	268	367	534
255488	BC063311	IBRDC2	264	1	264
255743	NM_198278	NPNT	117	374	490
255809	XM_172995		105	15	119
256471	BC029503	MGC33302	54	362	415
256536	BC042951	TCERG1L	116	411	526
256536	BC042951	TCERG1L	282	24	305
256710	BC014603	MGC26856	210	24	233
257044	BC032659	C1orf101	133	700	832
257062	BC043005	TMEM146	125	1	125
257313	NM_198152	UTS2D	92	28	119
259215	NM_001003693	C6orf21	282	16	297
259215	NM_001003693	C6orf21	119	16	134
259239	BC062670	WFDC11	62	26	87
260436	BC062213	C4orf7	68	18	85
261729	NM_152999	STEAP2	209	1	209

EntrezGeneID	Accession Transcript	Symbol	Length_AA	Nterm_AA	Cterm_AA
280664	NM_172006	WFDC10B	50	24	73
283238	BC034394	MGC34821	140	38	177
283298	NM_198474	OLFML1	255	143	397
283298	NM_198474	OLFML1	376	27	402
283375	BC027684	SLC39A5	196	19	214
283420	NM_207345	CLEC9A	114	120	233
283420	NM_207345	CLEC9A	79	112	190
283537	BC068556	LOC283537	61	17	77
283820	BC041131	NOMO2	185	31	215
283870	BC039154	MGC21830	154	50	203
283874	NM_001012731	LOC283874	144	313	456
283897	NM_175900	C16orf54	224	1	224
283971	NM_173619	MGC34761	155	27	181
283971	NM_173619	MGC34761	131	316	446
284013	NM_182566	VMO1	178	25	202
284099	BC034672	C17orf78	275	1	275
284129	NM_173626	SLC26A11	51	1	51
284186	NM_176520	TMEM105	80	50	129
284207	BC062252	METRNL	267	45	311
284266	NM_213602	CD33L3	309	20	328
284340	NM_198477	UNQ473	61	50	110
284348	NM_001031749	LYPD5	208	1	208
284369	BC034769	IZUMO1	329	22	350
284381	BC035001	LOC284381	121	28	148
284389	NM_173535	FLJ40235	81	20	100
284402	BC093509	LOC284402	74	23	96
284415	NM_196481	UNQ3033	118	17	134
284581	NM_001010882		104	1	104
284723	BC027998	SLC25A34	77	128	204
284759	XM_209363	SIRPB2	161	90	250
284996	BC019355	RNF149	178	223	400
285195	BC035779	SLC9A9	156	490	645
285203	BC060867	C3orf64	426	18	443
285313	NM_178822	IGSF10	301	2300	2600
285313	NM_178822	IGSF10	301	2300	2600
285368	NM_207351	PRRT3	445	28	472
285533	BC034385	RNF175	108	195	302
285613	BC063469	C5orf16	271	33	303
285613	BC063469	C5orf16	121	183	303
285754	NM_173671	FLJ37396	327	250	576
285761	BC035671	DCBLD1	165	248	412
286006	NM_182597	FLJ39575	51	5	55
286133	BC033153	SCARA5	318	83	400
286140	XM_209913		116	1	116
286256	BC041166	LCN12	337	19	355
286334	NM_001040063	LOC286334	92	16	107
286530	BC043610	P2RY6	102	258	359
338094	BC020674	C1orf179	126	460	585
338328	BC035610	LOC338328	116	50	165
338376	NM_176891	IFNE1	179	30	208
338811	BC028403	FAM19A2	102	30	131
338821	NM_001009562	LST-3TM12	89	119	207
338872	NM_178540	C1QTNF9	318	16	333
339168	BC107110	TMEM95	161	16	176
339168	BC107110	TMEM95	130	16	145
339366	NM_213604	ADAMTSL5	141	250	390
339804	XM_291016		89	27	115
339977	XM_291099	LOC339977	201	150	350
340146	BC067842	SLC35D3	114	303	416
340204	XM_929785	LOC340204	99	23	121
340273	NM_178559	ABCB5	245	1	245
340307	XM_496461		278	500	777
340547	BC043216	VSIG1	212	22	233
342510	NM_181449	CD300E	156	17	172
342865	XM_292785	LOC342865	123	248	370
343413	NM_001004310	FCRL6	201	50	250
345757	BC027332	UNQ1912	151	40	190
346007	NM_196283	EGFL11	191	350	540
346689	NM_196508	FLJ44186	260	1	260

EntrezGeneID	Accession Transcript	Symbol	Length_AA	Nterm_AA	Cterm_AA
347487	NM_001013403	RP11-35F15.2	291	71	361
347730	BC045113	LRRTM1	394	34	427
347902	BC047595	AMIGO2	359	39	397
348174	BC078143	LOC348174	155	27	181
348174	BC078143	LOC348174	131	316	446
348932	BC056757	SLC6A18	86	312	399
349152	NM_182634	FLJ36166	76	75	152
374383	NM_001009913	DKFZp686O24166	151	50	200
374395	BC051355	LOC374395	99	121	219
374768	BC033862	C17orf83	258	52	309
374768	BC033862	C17orf83	151	52	202
374768	BC033862	C17orf83	309	1	309
374819	XM_496238	FLJ34306	201	1	201
374882	BC064948	UNQ501	55	107	161
375387	BC044233	LRRC33	572	21	592
375387	BC044233	LRRC33	148	82	229
375567	NM_198570	UNQ739	298	28	325
375775	NM_152286	C9orf111	213	300	512
386724	NM_198722	AMIGO3	364	19	382
387070	OTTHUMT00000039675	C6orf66	139	1	139
387079	OTTHUMT00000042728	C6orf98	312	23	334
387597	NM_199351	C1orf32	167	20	186
387758	BC026873	LOC387758	194	18	211
387837	NM_205852	CLEC12B	172	61	232
387911	NM_001007537	LOC387911	234	100	333
388325	NM_207103	UNQ5783	103	43	145
388335	NM_001004313	LOC388335	79	82	160
388389	NM_213607	CCDC103	242	1	242
388394	BC033942	RPRML	56	1	56
388512	NM_207390	FLJ45910	113	194	306
388633	NM_001010978	LDLRAD1	144	62	205
388730	BC061592	TMEM81	193	31	223
388799	BC105792	RP5-1153D9.3	152	20	171
389012	NM_207403		144	30	173
389137	XM_371655	LOC389137	265	20	284
389558	NM_205855	UNQ1940	101	50	150
389730	XM_372094		179	61	239
389734	XM_372097	CNTNAP3B	336	230	565
389763	NM_001001670	FLJ46321	225	1352	1576
389850	XM_372205	LOC389850	270	54	323
390243	XM_372428	LOC390243	225	19	243
390928	NM_001004318	FLJ16165	285	134	418
391123	NM_001013661	VSIG8	244	21	264
392617	XM_374386	LOC392617	391	27	417
399716	XM_374767		101	82	182
399888	XM_374880	LOC399888	101	124	224
399947	BC068577	LOC399947	61	1	61
399948	NM_207429	FLJ45803	124	1	124
399979	BC031620	SNX19	201	100	300
400464	NM_001013670	LOC400464	231	1	231
400508	XM_375307	FLJ41766	61	49	109
400728	XM_927098	FAM87B	109	51	159
400943	NM_207480	UNQ5830	76	20	95
401152	BC017399	LOC401152	66	1	66
401278	NM_207500	FLJ44955	102	100	201
401507	NM_001012278	LOC401507	127	1	127
401612	XM_377034	MCART6	307	1	307
402604	XM_379939		213	23	235
414193	OTTHUMT00000047055	FAM23B	163	112	274
414196	OTTHUMT00000046761	C1Dorf31	468	23	490
414196	OTTHUMT00000046761	C1Dorf31	162	23	184
414308	NM_001009567	MRC1L1	121	22	142
414308	NM_001009567	MRC1L1	265	205	469
414767	NM_001012717		117	1	117
414919	BC064373	MGC70857	196	21	216
431705	NM_001002036	ASTL	192	92	283
439938	XM_498460		101	21	121
440163	BC044831	RNASE13	137	20	156
440699	NM_001005214	LRRC52	221	24	244

EntrezGeneID	Accession Transcript	Symbol	Length_AA	Nterm_AA	Cterm_AA
440764	XM_496476		54	251	304
440786	XM_496486	LOC440786	86	21	106
440871	XM_496556	LOC440871	131	21	151
440955	NM_001008269	TMEM89	137	23	159
441140	NM_001004349	FLJ45422	117	1	117
441402	XM_497024		373	26	398
441617	XM_497310	LOC441617	122	1	122
441631	XM_497334	TSPAN11	106	116	221
442117	BC047551	GALNT17	113	31	143
442760	XM_499891		202	61	262
444862	NM_001002923	IGFL4	106	19	124
493869	BC029424	LOC493869	109	47	155
497190	NM_001011680	LOC497190	136	181	316
548645	BC048318	bA16L21.2.1	76	168	243
619373	XM_940502	OACT4	201	26	226
619518	NM_001034847	C1orf191	83	18	100
641364	NM_001037234	TMEM75	55	84	138
641831	XM_935747	LOC641831	115	18	132
641928	NM_001039756	FLJ16734	98	23	120
642090	XM_942734	LOC642090	129	21	149
642132	XM_936279	LOC642132	86	525	610
642132	XM_936279	LOC642132	192	29	220
642149	XM_936295	LOC642149	260	1	260
642149	XM_936295	LOC642149	164	1	164
642253	XM_925796	LOC642253	56	42	97
642265	NM_001040065	RP11-95H8.1	137	1211	1347
642312	XM_925938	LOC642312	88	269	356
642373	XM_372097	LOC642373	131	368	498
642373	XM_372097	LOC642373	107	262	368
642564	XM_926048	LOC642564	120	24	143
642968	XM_926352	LOC642968	141	26	166
643664	XM_926969	LOC643664	103	1	103
643750	XM_927040	LOC643750	141	26	166
643792	XM_927073	FLJ37512	131	49	179
643797	XM_927076	LOC643797	105	23	127
643853	NM_001039770	FLJ45032	269	185	453
643904	XM_927169	LOC643904	193	1	193
643930	XM_930114	LOC643930	71	23	93
643940	XM_927199	LOC643940	93	1	93
644371	XM_929645	LOC644371	79	21	99
644571	XM_927686	FLJ16734	98	23	120
644975	NM_001039906	FLJ30064	115	18	132
645238	XM_930310	LOC645238	94	21	114
645294	XM_928339	LOC645294	79	1	79
645426	XM_928466	LOC645426	310	1	310
645460	XM_928492	LOC645460	100	23	122
645509	XM_928533	LOC645509	90	1	90
646100	XM_929061	LOC646100	92	39	130
646962	NM_001039792	UNQ338	85	31	115
647179	XM_930210	LOC647179	109	19	127
647291	NM_001039795		115	30	144
648629	XM_937698	LOC648629	57	167	223
648852	XM_940430	LOC648852	184	17	200
649891	XM_938970	LOC649891	128	22	149
649986	XM_939071	LOC649986	271	1	271
649986	XM_939071	LOC649986	169	83	271
649986	XM_939071	LOC649986	137	124	260
652222	XM_941605	LOC652222	76	1	76
652222	XM_941605	LOC652222	53	1	53
652626	XM_942172	LOC652626	445	16	460
652626	XM_942172	LOC652626	59	42	100
652674	XM_942255	LOC652674	266	263	528
652710	XM_942326	LOC652710	158	143	300
652760	XM_942393	LOC652760	71	22	92
652900	XM_942628	LOC652900	60	3	62
652900	XM_942628	LOC652900	155	1	155
652900	XM_942628	LOC652900	58	69	126
653141	XM_926169	LOC653141	224	1	224
653363	XM_927078	LOC653363	335	1	335

EntrezGeneID	Accession Transcript	Symbol	Length_AA	Nterm_AA	Cterm_AA
653363	XM_927076	LOC653363	410	1	410
653370	XM_930180	LOC653370	138	18	155
653370	XM_930180	LOC653370	138	18	155
653423	XM_929074	LOC653423	51	25	75
653486	XM_927639	LOC653486	73	23	95
653560	XM_928102	LOC653560	133	100	232
653600	XM_928349	LOC653600	76	26	101
653659	XM_930412	LOC653659	376	1	376
653659	XM_930412	LOC653659	297	1	297
653659	XM_930412	LOC653659	110	1	110
654055	XM_938478	LOC654055	87	1	87
654429	NM_001039029	LRTM2	151	35	185
	ENST00000340363	Q6UX52_HUMAN	246	20	265
	ENST00000340363	Q6UX52_HUMAN	124	70	193
	ENST00000357601	P11388-3	284	465	749
	ENST00000357601	P11388-3	254	1219	1472
	ENST00000300671	O00431_HUMAN	101	1	101
	ENST00000375352	Q6UXZ3_HUMAN	177	18	194
	ENST00000375352	Q6UXZ3_HUMAN	148	18	165
	ENST00000360372	Q6ZRB7_HUMAN	174	16	189
	ENST00000310542	Q9HBS9_HUMAN	131	17	147
	ENST00000338989	NR_002807.1	66	31	96
	ENST00000359701	Q8NAJ9_HUMAN	132	1	132
	ENST00000378074	Q6XRZ0_HUMAN	77	1	77
	ENST00000382556	Q6XYA8_HUMAN	95	1	95
	ENST00000330146	Q8N852_HUMAN	109	51	159
	ENST00000382566	Q6ZTC8_HUMAN	88	1	88
	ENST00000341462	Q7Z2S2_HUMAN	59	47	105
	OTTHUMT00000045396	RP11-395N17.1	154	17	170
	OTTHUMT00000044566	RP11-412K4.1	212	21	232
	OTTHUMT00000044734	RP11-478K15.4	85	1	85
	OTTHUMT00000047374	RP11-48B24.3	134	20	153
	OTTHUMT00000044610	RP11-50D16.2	138	20	157
	OTTHUMT00000044616	RP11-50D16.3	324	24	347
	OTTHUMT00000046596	RP11-522H2.2	524	22	545
	OTTHUMT00000048596	RP11-522H2.2	201	300	500
	OTTHUMT00000047966	RP11-523O18.2	156	23	178
	OTTHUMT00000053989	C9orf56	80	1	80
	OTTHUMT00000041352	RP1-223E3.1	399	16	414
	OTTHUMT00000041352	RP1-223E3.1	121	17	137
	OTTHUMT00000043538	RP1-238O23.3	297	15	311
	OTTHUMT00000043538	RP1-238O23.3	146	15	160
	ENST00000355850	Q71RG6_HUMAN	186	23	208
	ENST00000382566	Q6ZTC8_HUMAN	123	1	123
	OTTHUMT00000056459	RP11-38O23.2	139	21	159
	OTTHUMT00000044162	RP11-45B20.2	315	19	333
	OTTHUMT00000044162	RP11-45B20.2	134	200	333
	OTTHUMT00000045820	RP11-480K16.1	172	24	195
	OTTHUMT00000047976	RP11-507P23.5	141	19	159
	OTTHUMT00000075379	RP1-151B14.4	202	39	240
	OTTHUMT00000250584	AC009333.1	291	1	291
	OTTHUMT00000250259	AC006026.7	105	36	140
	OTTHUMT00000246870	AC007321.4	156	119	274
	OTTHUMT00000133879	AC023356.2	268	1	268
	OTTHUMT00000133062	AC004775.3	205	345	549
	OTTHUMT00000129787	FAM95B	128	30	157
	OTTHUMT00000102340	AC068580.3	57	1	57
	OTTHUMT00000096863	OR2AJ1	60	36	95
	OTTHUMT00000087736	RP11-212H11.4	97	25	121
	OTTHUMT00000083793	RP11-541H12.2	222	23	244
	OTTHUMT00000078312	C1QR	125	182	306
	OTTHUMT00000078312	C1QR	160	23	182
	OTTHUMT00000075544	RP3-515N1.6	130	1	130
	OTTHUMT00000075359	AC006946.3	83	339	421
	OTTHUMT00000075223	CTA-747E2.6	133	139	271
	OTTHUMT00000075114	RP3-515N1.2	149	21	169
	OTTHUMT00000074704	AF111168.1	242	334	575
	OTTHUMT00000074704	AF111168.1	303	32	334
	OTTHUMT00000073960	AL133453.1	156	1	156

EntrezGeneID	Accession Transcript	Symbol	Length_AA	Nterm_AA	Cterm_AA
	OTTHUMT00000072425	AL359220.2	266	52	317
	OTTHUMT00000072114	AL049838.2	276	499	774
	OTTHUMT00000072114	AL049838.2	472	28	499
	OTTHUMT00000071941	AL161751.1	198	201	398
	OTTHUMT00000059884	HCA112	56	137	192
	OTTHUMT00000059846	LOC168433	165	212	376
	OTTHUMT00000059677	CAS1	58	1	58
	OTTHUMT00000059559	mbxx_chr7.142.004.a	52	73	124
	OTTHUMT00000059437	FLJ11000	47	35	81
	OTTHUMT00000059326	LOC168391	121	323	443
	OTTHUMT00000059291	mbxx_ts.110.002.a	72	22	93
	OTTHUMT00000059291	mbxx_ts.110.002.a	266	93	358
	OTTHUMT00000059291	mbxx_ts.110.002.a	292	335	626
	OTTHUMT00000078312	C1QR	278	306	583
	OTTHUMT00000075712	CTA-964G1.2	156	33	188
	OTTHUMT00000073352	AL591771.1	212	203	414
	OTTHUMT00000073352	AL591771.1	182	414	595
	OTTHUMT00000060259	MGC5442	156	35	190
	OTTHUMT00000059502	FLJ13576	127	300	426
	OTTHUMT00000059326	LOC168391	296	28	323
	OTTHUMT00000059273	mbxx_chr7.2.007.a	130	27	156
	OTTHUMT00000059273	mbxx_chr7.2.007.a	237	156	392
	OTTHUMT00000059204	LR8	65	1	65
	OTTHUMT00000059181	mbhmm_H_NH0296A10_F115294.fgenesh2_4	329	207	535
	OTTHUMT00000059181	mbhmm_H_NH0296A10_F115294.fgenesh2_4	343	1158	1500
	OTTHUMT00000058091	RP3-525N14.6	109	112	220
	OTTHUMT00000056919	RP11-479E16.1	185	35	219
	OTTHUMT00000056919	RP11-479E16.1	302	231	532
	OTTHUMT00000055102	C9orf136--suspended	383	49	431
	OTTHUMT00000055102	C9orf136--suspended	152	49	200
	OTTHUMT00000052325	RP11-331F9.6	201	63	263
	OTTHUMT00000050565	RP11-129M16.2	343	24	366
	OTTHUMT00000050565	RP11-129M16.2	396	446	841
	OTTHUMT00000050565	RP11-129M16.2	268	887	1154
	OTTHUMT00000050507	RP11-537G20.1	273	1	273
	OTTHUMT00000050507	RP11-537G20.1	389	273	661
	OTTHUMT00000050507	RP11-537G20.1	218	661	878
	OTTHUMT00000050075	RP11-18I14.8	51	84	134
	OTTHUMT00000049610	RP11-34E5.1	195	28	222
	OTTHUMT00000049030	RP11-369J21.6	98	26	123
	OTTHUMT00000049030	RP11-369J21.6	57	36	92
	OTTHUMT00000047248	RP11-59G22.1	395	23	417
	OTTHUMT00000047248	RP11-59G22.1	218	636	853
	OTTHUMT00000045764	RP11-90L1.3	121	30	150
	OTTHUMT00000045646	RP11-113J24.1	129	1	129
	OTTHUMT00000045646	RP11-113J24.1	282	479	760
	OTTHUMT00000078607	RP5-836N17.2	262	1	262
	OTTHUMT00000050853	RP11-391M7.1	125	632	756
	OTTHUMT00000050853	RP11-391M7.1	608	25	632
	OTTHUMT00000049081	RP11-137H2.2	119	114	232
	OTTHUMT00000045963	RP11-199F6.1	109	218	326
	OTTHUMT00000044528	RP11-421P11.2	220	31	250
	OTTHUMT00000044528	RP11-421P11.2	75	250	324
	OTTHUMT00000042062	RP1-84N20.1	149	1	149
	OTTHUMT00000041956	RP3-493F7.3	191	119	309
	OTTHUMT00000041771	RP1-249I4.1	65	1	65
	OTTHUMT00000072630	AB019441.12	147	366	512
	OTTHUMT00000072560	AC005280.1	97	29	125
	OTTHUMT00000072301	AL357153.2	328	1	328
	OTTHUMT00000059789	mbhmm_gw12844788.99.40.1.5e-25.gw_1	220	1	220
	OTTHUMT00000048137	RP11-179B15.3	108	1	108
	OTTHUMT00000040275	RP3-329A5.4	169	29	197
	OTTHUMT00000042149	RP3-403A15.3	119	800	918
	ENST00000324446	CY014_HUMAN	113	1	113
	ENST00000314117	Q8IUW5_HUMAN	244	28	271
	ENST00000333833	Q8N2L8_HUMAN	130	1	130
	ENST00000328411	Q8IV56_HUMAN	135	1	135
	ENST00000332749	Q9H5Q3_HUMAN	168	1	168
	ENST00000341675	XR_015253.1	377	541	917
	ENST00000356867	Q9HAJ0_HUMAN	119	1	119

EntrezGeneID	Accession Transcript	Symbol	Length_AA	Nterm_AA	Cterm_AA
	ENST00000322282	GRAMD1B	163	1	163
	ENST00000216241	Q5JY13_HUMAN	98	665	762
	ENST00000238971	Q9H354_HUMAN	126	1	126
	ENST00000319163		67	22	88
	ENST00000283694		102	19	120
	ENST00000298350	Q96JQ7_HUMAN	64	25	88
	ENST00000299612	Q8NH75_HUMAN	220	33	252
	ENST00000330155	Q8N0W1_HUMAN	139	49	187
	ENST00000329282	Q96P86_HUMAN	74	1	74
	ENST00000343152	Q6ZU86_HUMAN	134	19	152
	ENST00000321460	DNAJC4	241	1	241
	ENST00000243152	TYRL	134	1	134
	ENST00000246222	Q9H4V5_HUMAN	262	18	279
	ENST00000356298		110	21	130
	OTTHUMT00000074924	RP3-355C18.1	95	283	377
	OTTHUMT00000039753	RP3-380B6.2	116	27	142
	OTTHUMT00000075010	RP3-402G11.11	362	182	543
	OTTHUMT00000040871	RP3-442L6.3	227	19	245
	OTTHUMT00000075517	RP3-526I14.3	134	23	156
	OTTHUMT00000074941	RP5-821D11.5	208	22	229
	OTTHUMT00000075112	XXbac-B444P24.1	248	226	473
	OTTHUMT00000072507	AC004846.1	116	186	301
	OTTHUMT00000072939	AL049780.1	155	37	191
	OTTHUMT00000074401	AL117259.4	286	24	309
	OTTHUMT00000059311	mbnmm_gw12844768.61.18.2.7e-26.gw_1	162	1	162
	OTTHUMT00000059202	mbxx_chr7.140.001.a	121	23	143
	OTTHUMT00000049891	RP11-108L7.10	275	30	304
	OTTHUMT00000050204	RP11-127L20.4	105	443	547
	OTTHUMT00000043214	RP11-160E12.4	162	30	191
	OTTHUMT00000043201	RP11-270C4__A.1	133	21	153
	OTTHUMT00000040965	RP11-345L23.2	181	172	352
	ENST00000259726	Q7Z560_HUMAN	265	16	280
	ENST00000383319	Q861E6_HUMAN	319	22	340
	ENST00000377425	NR_002823.1	85	1	85
	ENST00000303432		155	18	172
	ENST00000305754	OR5E1P	86	1	86
	ENST00000311003		319	22	340
	ENST00000312214		209	24	232
	ENST00000311755	HIGD2BP	106	1	106
	ENST00000310146	CR030_HUMAN	84	482	565
	ENST00000318487		274	1	274
	ENST00000342748		139	18	156
	OTTHUMT00000075543	RP3-412A9.5	134	19	152
	OTTHUMT00000079861	RP4-565L9.2	112	24	135
	OTTHUMT00000051753	C9orf145	218	29	246
	ENST00000323595		101	20	120
	ENST00000316397		95	30	124
	ENST00000313957	Q8N9G5_HUMAN	368	17	384
	ENST00000314747		170	38	207
	ENST00000318969	XR_001004.1	145	25	169
	ENST00000324982	Q9NT46_HUMAN	121	23	143
	ENST00000315606		64	20	83
	ENST00000316575	Q8N409_HUMAN	472	1	472
	ENST00000327725	Q9H5Q9_HUMAN	105	41	145
	ENST00000331747		97	153	249
	ENST00000329738		142	24	165
	ENST00000334994	Q7Z2Q7_HUMAN	161	348	508
	ENST00000317346	Q9H374_HUMAN	64	20	83
	ENST00000338640	Q4G193_HUMAN	74	19	92
	ENST00000340783		300	30	329
	ENST00000345013	Q8NH71_HUMAN	141	1	141
	ENST00000341508		144	35	178
	ENST00000333145	Q7Z3M5_HUMAN	132	16	147
	ENST00000344740		67	22	88
	ENST00000357027	Q96NJ4_HUMAN	124	16	139
	ENST00000355821		67	22	88
	ENST00000358533		168	1	168
	ENST00000328274		223	1	223
	ENST00000358379		84	26	109

EntrezGeneID	Accession Transcript	Symbol	Length_AA	Nterm_AA	Cterm_AA
	ENST00000357453		100	20	119
	ENST00000355529		86	1	86
	ENST00000360964		83	41	123
	ENST00000354502		67	22	88
	ENST00000374099	Q9H374_HUMAN	64	20	83
	ENST00000374774	Q6UXS3_HUMAN	91	18	108
	ENST00000374946		73	91	163
	ENST00000375181	Q6ZRJ6_HUMAN	58	28	85
	ENST00000375211	Q6UWG9_HUMAN	64	26	89
	ENST00000375868	Q6ZVM5_HUMAN	250	1	250
	ENST00000376604	Q6ZR28_HUMAN	144	27	170
	ENST00000377210	Q6ZST7_HUMAN	204	1	204
	ENST00000377230		100	20	119
	ENST00000312946		102	19	120
	ENST00000377273		372	18	389
	ENST00000377712	Q6NT18_HUMAN	101	43	143
	ENST00000378005	Q6ZP42_HUMAN	116	19	134
	ENST00000378296		132	1	132
	ENST00000340623	Q8NBX4_HUMAN	108	20	127
	ENST00000342204	Q6UXU0_HUMAN	104	34	137
	ENST00000358718		71	18	88
	ENST00000369468	Q6ZUB6_HUMAN	187	1	187
	ENST00000377953	U655_HUMAN	71	23	93
	ENST00000378367	Q6ZVP5_HUMAN	138	1	138
	ENST00000378490	Q6ZT81_HUMAN	226	1	226
	ENST00000258173		160	34	193
	ENST00000379712	Q6ZVR2_HUMAN	114	16	129
	ENST00000379793	Q6ZN60_HUMAN	113	17	129
	ENST00000360029		90	1	90
	ENST00000360065	Q6ZU34_HUMAN	100	34	133
	ENST00000360387	Q6ZUE4_HUMAN	125	1	125
	ENST00000360819	Q71MF9_HUMAN	94	1	94
	ENST00000361002		103	27	129
	ENST00000361013		200	1	200
	ENST00000361068	Q6ZV15_HUMAN	152	16	167
	ENST00000343736		221	1	221
	ENST00000362196	Q6ZSV2_HUMAN	104	27	130
	ENST00000362204	Q6UXR8_HUMAN	90	33	122
	ENST00000362662	Q6ZSP9_HUMAN	129	1	129
	ENST00000362776		76	22	97
	ENST00000362854	Q6ZWG3_HUMAN	122	21	142
	ENST00000362898	Q6ZS21_HUMAN	114	20	133
	ENST00000363494	Q6ZU40_HUMAN	115	30	144
	ENST00000363509	Q6ZRD9_HUMAN	122	21	142
	ENST00000363649		83	41	123
	ENST00000363727	Q6ZRW1_HUMAN	120	35	154
	ENST00000363764	XR_000649.1	135	34	168
	OTTHUMT00000006282	C1orf134	64	20	83
	OTTHUMT00000006325	RP11-56N19.1	315	35	349
	OTTHUMT00000007658	RP5-1056L3.6	268	367	634
	OTTHUMT00000010655	RP11-490K7.2	65	85	149
	OTTHUMT00000011175	RP11-460I13.3	93	68	160
	OTTHUMT00000021931	RP11-8J9.3	402	94	495
	OTTHUMT00000023027	RP5-997D24.4	83	18	100
	OTTHUMT00000026823	RP11-76L19.1	179	62	240
	OTTHUMT00000026931	RP11-181C21.1	203	94	296
	OTTHUMT00000027683	RP11-12C17.1	210	1	210
	OTTHUMT00000029677	RP4-684L20.3	351	22	372
	OTTHUMT00000029699	RP4-539F20.2	192	25	216
	OTTHUMT00000030046	RP4-716F6.2	378	51	428
	OTTHUMT00000036919	FAM75A3	137	1211	1347
	OTTHUMT00000039980	RP1-130G2.2	248	134	381
	OTTHUMT00000055837	RP13-928P6.1	260	1	260
	ENST00000379466	Q6UWF6_HUMAN	96	18	113
	ENST00000381273		262	18	279
	ENST00000362053	Q6ZNQ8_HUMAN	92	32	123
	ENST00000342371		67	22	88
	ENST00000383354		60	19	78
	ENST00000383420	Q5SQ64_HUMAN	219	16	234

EntrezGeneID	Accession Transcript	Symbol	Length_AA	Nterm_AA	Cterm_AA
	OTTHUMT00000009659	RP3-469D22.2	489	176	664
	OTTHUMT00000056950	CXorf26	410	1	410
	OTTHUMT00000057301	RP11-217H1.1	156	30	185
	OTTHUMT00000046277	RP11-58P3.2	118	302	419
	OTTHUMT00000060000	AF56CDS3	175	1	175
	OTTHUMT00000075029	RP1-257I20.4	107	1	107
	OTTHUMT00000075533	RP3-508I15.4	482	236	717
	OTTHUMT00000075840	XXbac-B562F10.3	208	43	250
	OTTHUMT00000090076	RP11-536L3.2	195	30	224
	ENST00000343465		176	1	176
	ENST00000361614	Q6NBL2_HUMAN	115	30	144
	ENST00000374345		128	26	153
	ENST00000295339		99	20	118
	ENST00000304627		115	18	132
	ENST00000326734		109	51	159
	ENST00000316239		242	20	261
	ENST00000345101		128	30	157
	ENST00000369196		141	579	719
	ENST00000355748		71	24	94
	ENST00000360844		138	1	138
	ENST00000358610		225	1	225
	ENST00000361129		64	27	90
	ENST00000361502		79	21	99
	ENST00000376283		100	23	122
	ENST00000316269		128	30	157
	ENST00000378661		86	1	86
	ENST00000381245		51	20	70
	OTTHUMT00000044518		107	1	107
	OTTHUMT00000050157		146	199	344
	ENST00000299978	TMEM63	74	62	135
	ENST00000284168		51	226	286
	ENST00000302079		194	17	210
	ENST00000333487		115	15	129
	ENST00000340301		259	1	259
	ENST00000345111		88	1	88
	OTTHUMT00000049007		96	39	134
	OTTHUMT00000055114		165	20	184
	OTTHUMT00000055617		238	31	268
	OTTHUMT00000059138		96	196	291
	OTTHUMT00000059144			0	164
	OTTHUMT00000059161		153	20	172
	OTTHUMT00000059164		153	20	172
	OTTHUMT00000059189		153	20	172
	OTTHUMT00000059226		100	83	182
	OTTHUMT00000059281		315	514	828
	OTTHUMT00000059325		263	24	286
	OTTHUMT00000059352		309	23	331
	OTTHUMT00000059540		181	1	181
	OTTHUMT00000059634		54	18	71
	OTTHUMT00000059655		167	559	725
	OTTHUMT00000059699		279	614	892
	OTTHUMT00000059740		245	16	261
	OTTHUMT00000059781		182	18	199
	OTTHUMT00000059847		155	20	174
	OTTHUMT00000059855		66	20	85
	OTTHUMT00000059867		128	20	147
	OTTHUMT00000059906		184	23	206
	OTTHUMT00000059945		552	23	574
	OTTHUMT00000059956		141	23	163
	OTTHUMT00000060063		105	20	124
	OTTHUMT00000060070		108	20	127
	OTTHUMT00000060082		417	21	437
	OTTHUMT00000050112		110	20	129
	OTTHUMT00000050115		110	20	129
	OTTHUMT00000060129		140	20	159
	OTTHUMT00000060132		110	20	129
	OTTHUMT00000073651		189	28	216
	OTTHUMT00000074058		233	204	436
	OTTHUMT00000075757		407	27	433

EntrezGeneID	Accession Transcript	Symbol	Length_AA	Nterm_AA	Cterm_AA
	OTTHUMT00000075879		175	25	199
	OTTHUMT00000059498		139	26	164
	OTTHUMT00000059779		367	18	384
	OTTHUMT00000059852		116	20	135
	OTTHUMT00000060058		140	20	159
	OTTHUMT00000060068		114	20	133
	OTTHUMT00000060094		95	23	117
	OTTHUMT00000060113		145	20	164
	OTTHUMT00000109071		181	30	210
	OTTHUMT00000109486		76	1	76
	OTTHUMT00000060133		137	18	154
	OTTHUMT00000060134		115	30	144
	OTTHUMT00000109520		131	1	131
	OTTHUMT00000132432		206	1	206
	OTTHUMT00000145099		148	18	165
	OTTHUMT00000146063		526	272	797
	OTTHUMT00000146120		387	1	387
	OTTHUMT00000146163		134	1	134
	OTTHUMT00000146199		157	586	742
	OTTHUMT00000146235		270	1	270
	OTTHUMT00000146251		467	26	492
	OTTHUMT00000146256		84	477	560
	OTTHUMT00000146270		365	1	365
	OTTHUMT00000146292		312	117	428
	OTTHUMT00000146398		365	19	383
	OTTHUMT00000146415	-	250	44	293
	OTTHUMT00000147346		368	30	397
	OTTHUMT00000147369		235	18	252
	OTTHUMT00000147382		321	21	341
	OTTHUMT00000147512		120	31	150
	OTTHUMT00000147591	-	481	20	500
	OTTHUMT00000147626		305	20	324
	OTTHUMT00000147871		295	31	325
	OTTHUMT00000147951		103	30	132
	OTTHUMT00000147958	-	126	23	148
	OTTHUMT00000147964		446	30	475
	OTTHUMT00000147992		164	1	164
	OTTHUMT00000148015		182	27	208
	OTTHUMT00000148035		173	1	173
	OTTHUMT00000148044		308	25	332
	OTTHUMT00000060139		132	20	151
	OTTHUMT00000060156		143	20	162
	OTTHUMT00000146213	-	447	20	466
	OTTHUMT00000150674		131	20	150
	OTTHUMT00000150747		594	37	630
	OTTHUMT00000150920	-	433	314	746
	OTTHUMT00000151031	-	327	1	327
	OTTHUMT00000151255		200	15	214
	OTTHUMT00000151300		211	1	211
	OTTHUMT00000151328		298	1	298
	OTTHUMT00000151349		114	26	139
	OTTHUMT00000151355		104	19	122
	OTTHUMT00000151359	-	87	23	109
	OTTHUMT00000151368		97	1	97
	OTTHUMT00000151376		92	1	92
	OTTHUMT00000151631		101	18	118
	OTTHUMT00000151633		142	16	157
	OTTHUMT00000151656		186	26	211
	OTTHUMT00000151665	-	68	28	95
	OTTHUMT00000151671		254	1	254
	OTTHUMT00000151682		307	30	336
	OTTHUMT00000151696		82	106	187
	OTTHUMT00000059418		172	1	172
	OTTHUMT00000060181	-	294	206	499
	OTTHUMT00000059484	-	113	49	161
	OTTHUMT00000151706	-	87	109	195
	OTTHUMT00000151723	-	307	22	328
	OTTHUMT00000151728		194	27	220
	OTTHUMT00000151781		136	51	186

EntrezGeneID	Accession Transcript	Symbol	Length_AA	Nterm_AA	Cterm_AA
	OTTHUMT00000151815	-	253	30	282
	OTTHUMT00000151837	-	486	35	520
	OTTHUMT00000151854		154	1	154
	OTTHUMT00000151873		175	35	209
	OTTHUMT00000151916		231	1	231
	OTTHUMT00000152014		172	545	716
	OTTHUMT00000151806	-	293	34	326
	OTTHUMT00000152155		246	43	288
	OTTHUMT00000152494	-	148	25	172
	OTTHUMT00000154843		87	22	108
	OTTHUMT00000154961	-	366	30	395
	OTTHUMT00000155207		83	1	83
	OTTHUMT00000155231		82	20	101
	OTTHUMT00000155233		84	20	103
	OTTHUMT00000155460		74	1	74
	OTTHUMT00000155530		89	25	113
	OTTHUMT00000155855		445	641	1085
	OTTHUMT00000156040		131	35	165
	OTTHUMT00000194073	-	139	1	139
	OTTHUMT00000252009	-	95	1	95
	OTTHUMT00000261461		128	30	157
	OTTHUMT00000261916		76	1	76
	OTTHUMT00000151710	-	212	19	230
	OTTHUMT00000151804	-	95	1	95
	OTTHUMT00000151915		157	21	177
	OTTHUMT00000155380		114	31	144
	OTTHUMT00000261964		211	23	233
	OTTHUMT00000262066	-	79	31	109
	OTTHUMT00000262071	-	368	469	836
	OTTHUMT00000262286	-	97	198	294
	OTTHUMT00000262736		142	1	142
	OTTHUMT00000262822		173	33	205
	OTTHUMT00000263188		673	25	697
	OTTHUMT00000263302		493	1	493
	OTTHUMT00000265335		99	1	99
	OTTHUMT00000265370		93	15	107
	OTTHUMT00000265800		224	21	244
	OTTHUMT00000266473		429	73	501
	OTTHUMT00000266872		131	35	165
	OTTHUMT00000270278		148	17	164
	OTTHUMT00000270334		282	17	298
	OTTHUMT00000271666		224	75	298
	OTTHUMT00000265229		84	18	101
	OTTHUMT00000265448	-	99	1	99
	OTTHUMT00000266226		471	59	529
	OTTHUMT00000266276	-	54	1	54
	OTTHUMT00000270362	-	132	26	157
	OTTHUMT00000270392		305	1	305
	OTTHUMT00000271651		266	1	266
	OTTHUMT00000271944		208	24	231
	OTTHUMT00000272035	-	297	1	297