**Dataset Brief**

**Title:** Exploring analytical proteomics platforms towards the definition of human Cardiac Stem Cells receptome

**Authors:** Patrícia Gomes-Alves[1,2], Margarida Serra [1,2], Catarina Brito[1,2], Luis R-Borlado[3], Juan A. López[4], Jesús Vázquez[4], Manuel J.T. Carrondo[1,2,5], António Bernad[6], Paula M. Alves[1,2,*]

**Affiliations:**

[1] iBET, Instituto de Biologia Experimental e Tecnológica, Apartado 12, 2781-901 Oeiras, Portugal;

[2] Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Av. da República, 2780-157 Oeiras, Portugal;

[3] Coretherapix, Tres-Cantos (Madrid), Spain;

[4] Laboratorio de Proteómica Cardiovascular and Unidad de Proteómica; Centro Nacional de Investigaciones Cardiovasculares, Madrid, Spain;

[5] Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Monte da Caparica, Portugal;

[6] Centro Nacional de Biotecnología, Madrid, Spain.

***Corresponding author:** Paula M. Alves, ITQB-UNL/iBET, Apartado 12, 2781-901 Oeiras, Portugal (Phone: +351 214469421; Fax:+351 214421161; e-mail: **marques@ibet.pt**; URL: **http://tca.itqb.unl.pt**)

**Abbreviations:** CM, cardiomyocytes; CVDs, Cardiovascular diseases; GF, growth factors; hCSC, Human Cardiac Stem Cells; IPA, Ingenuity Pathway Analysis software; MCX, mixed-mode cationic exchange; TMDs, transmembrane domains.

**Keywords:** Human Cardiac Stem Cells; nanoLC-MS/MS; Pre-fractionation; Proteomics; Receptome

**Total number of words**: 2938 (excl. the title page and legends for the supplemental material).

**Abstract**

Human cardiac stem cells (hCSC) express a portfolio of plasma membrane receptors that are involved in the regulatory auto/paracrine feedback loop mechanism of activation of these cells, and consequently contribute to myocardial regeneration. In order to attain a comprehensive description of hCSC receptome and overcoming the inability demonstrated by other technologies applied in receptor identification, mainly due to the transmembrane nature, high hydrophobic character and relative low concentration of these proteins, we have exploited and improved a proteomics workflow. This approach was based on the enrichment of hCSC plasma membrane fraction and addition of pre-fractionation steps prior to MS analysis. More than one hundred plasma membrane receptors were identified. The data reported herein constitute a valuable source of information to further understand cardiac stem cells activation mechanisms and the subsequent cardiac repair process. Data have been deposited to the ProteomeXchange with identifier PXD001117

Cardiovascular diseases (CVDs) are the leading cause of death worldwide, namely ischaemic heart disease that was responsible for the highest number of deaths during the past decade according to the World Health Organization (WHO). Moreover, WHO estimates that CVDs will remain the single leading cause of death, reaching about 23 million deaths by 2030 [1]. Particularly, acute myocardial infarction (AMI) severely affects patients' heart muscle and microvasculature, critically decreasing the number of functional cardiomyocytes (CM). Stem cell and protein based therapies are promising therapeutic strategies for cardiac repair since it was found that the adult myocardium homes a minor population of resident cardiac stem cells (CSC) capable to contribute to contractile myocardium regeneration [2-4]. The surviving CM under pathological stress produce a set of growth factors (GF) and cytokines that promote CM survival and activate endogenous CSC niches. Consequently, an auto/paracrine loop is triggered to maintain the GF production, which stimulates CSC activation and differentiation into new CM, endothelial and smooth muscle cells able to repair the damaged myocardium [5, 6].

Although CSC activation is responsible for a continuous, but low level, repopulation of the myocardium, it is not robust and durable enough to have significant beneficial physiological/anatomical impact in severe and acute myocardial losses, as occurs in most AMI

episodes [7]. However, local administration of GF has been shown to be efficient in activating CSC, improving cardiac output post-AMI through the formation of new vascularized and functional autologous myocardium. Additionally, it was also demonstrated in animal models that GF administration is a very effective way to promote the regeneration of myocardium after ischemia [6, 8, 9]. Therefore, it is of outmost importance to define the GF receptors involved in the paracrine-autocrine feedback loop between CM and hCSC, to disclose the signaling pathways beneath myocardial regeneration.

In this study we investigated the profile of proteins expressed in hCSC, namely at the receptome level, using an optimized proteomics pipeline. Proteomics methodologies now available enable the high-throughput identification of almost entire proteomes and sub-proteomes of interest and assessment of the complexity of proteins present in different cells, unveiling their potential biological roles. Moreover, technological advances, mainly in mass spectrometry (MS) instrumentation, resulted in a substantial increase in plasma membrane proteins representation in large data sets [10-12]. Other attempts for partial hCSC receptome characterization using specific proteomics techniques have been previously reported. Moscoso *et al.* described an approach based on biotinylation of plasma membrane proteins where only 36 proteins were identified [13]. Moreover, most of the proteins listed are just plasma membrane associated or contained a single transmembrane domain and only a few receptors were identified. To overcome this inability to efficiently identify receptors, we have integrated different protocols and approaches into a workflow based on optimized plasma membrane protein fraction preparation, pre-fractionation and state-of-the-art MS.

**Enriched plasma membrane fractions:** hCSC isolation and cell characterization were previously reported [13, 14]. After expansion, cells were harvested (passage 6/7) by scraping and washed twice with PBS. Cell pellets ($5x10^6$ cells/pellet) were resuspended in lysis buffer and lysed with 30 passes through the 301/2 Gauge needle, at 4ºC. Enriched plasma membrane fractions were obtained as described by Dornemeyer *et al*. [12], and finally rinsed twice with cold water at 20,000 x g, 30 min at 4 ºC (detailed protocol described in supporting information document 1). Obtained fractions were separated in a SDS-PAGE gel (~25 µg/lane) and the presence of protein markers for other subcellular compartments/intracellular membranes was detected by Western-blot (supporting information document 2).

**1D-MALDI-TOF/TOF**: To proceed with hCSC receptome characterization, an in-gel proteomics approach was initially tested. An entire gel lane (~50 µg per lane) was excised and sliced in 60 bands that were independently digested (trypsin, Promega), desalted and concentrated sequentially in custom-made chromatographic microcolumns with increasing hydrophobic material (Poros R2, Poros R3 and graphite powder) [15]. The flow-throughs were directly spotted on a MALDI plate using 0.6 µL of 5 mg/mL α-Cyano-4-hydroxycinnamic acid (Sigma) and analyzed on a 4800 Plus MALDI-TOF/TOF analyser using 4000 Series Explorer Software v.3.5.3 (ABSciex). The top 15 precursor ions from the MS spectra were selected for MS/MS analysis. The MS and MS/MS data were analyzed in combined search mode (MS+MS/MS) using GPS Explorer Software v.3.6 (ABSciex) and MASCOT. The search parameters considered were: monoisotopic peptide mass values, maximum precursor mass tolerance of 50 ppm and a maximum fragment mass tolerance of 0.3 Da. The searches were performed against SwissProt database (*Homo sapiens*, 20319 sequences). One missed cleavage was allowed; carboxamidemethylation of cysteines and oxidation of methionines were set as fixed and variable modifications, respectively. Protein identification was only accepted when significant protein homology scores were obtained ($p<0.05$) and at least one peptide was fragmented with a significant individual ion score ($p<0.05$). More than 120 unique proteins were identified (Supporting Information Table ST1) using this approach, only the more abundant proteins in each band were identified. In order to access the hydrophobicity and number of transmembrane domains of the identified proteins, bioinformatics tools were used to calculate putative transmembrane domains (TMD) and GRAVY index (http://www.cbs.dtu.dk/services/TMHMM/ and http://www.bioinformatics.org/sms2/protein_gravy.html). The number of transmembrane proteins was still very small (43 proteins), and from those only 3 were classified as receptors (Figure 1). While the enrichment in membrane proteins is demonstrated, a large contamination with proteins from other organelles is also observed. From this analysis and in accordance with previous reports in the literature [16] we confirmed that this was not the most effective strategy to identify less abundant proteins, and specifically those with more transmembrane domains/receptors.

**nanoLC-LTQ-Orbitrap**: To improve the numbers of receptors accessed we have proceeded with a different approach comprising a nanoLC fractionation previous to MS analysis. hCSCs membrane enriched fractions were prepared as described above and loaded (~50 µg) on a SDS-PAGE gel for cleaning and trypsin digestion, as described [17]. Proteins were concentrated in a single gel band that

was visualized by colloidal-Coomassie staining, excised, and cut in 15-20 small pieces. Sample was subjected to reduction and alkylation, and digested with trypsin. TFA was added to a final concentration of 1% and the peptides were desalted onto HLB Oasis cartridges and dried-down. Sample tryptic peptides (3 µg per injection) were analyzed by LC-MS, using the LTQ-Orbitrap XL (Thermo Scientific). Proteins were identified with Proteome Discoverer 1.3 (Thermo Scientific); two mixed cleavages were allowed, carboxamidemethylation of cysteines and oxidation of methionines were set as fixed and variable modifications, respectively, and an error of 20 ppm or 1.2 Da was set for full MS or MS/MS spectra searches, respectively. Peptide identifications were validated using the probability ratio and the refined FDR method [18, 19]. Peptide identifications were firstly filtered out using 0.05 FDR prior to the application of subsequent, more restrictive criteria (see below). Protein classification was performed using Ingenuity Pathway Analysis software (IPA). GRAVY indexes and TMDs were investigated for these proteins as described above (Figure 1). A list of more than 600 proteins was obtained, including 218 proteins with TMDs and 23 receptors (e.g. IGF2R, ABCC1 …) (Figure 1; Supporting Information Table ST2). Eighty of the proteins identified were also found in our MALDI-TOF/TOF experiment, and 18 were already described in Moscoso *et al*. study [13]. CD71, MMP-14, and a few others were consistently identified. Proteins described to be involved in the cardiac function (sarcomeric proteins) were also found. Clear improvements in hCSCs receptome analysis protocol were accomplished, although we pursued a more complete dataset covering proteins with more TMDs and better representation of receptors.

**MCX-nanoLC-LTQ-Orbitrap**: Addition of an extra pre-fractionation step prior to nanoLC was tested. Enriched plasma membrane fractions where digested as above and the resulting peptides further fractionated through an offline mixed-mode cationic exchange (MCX) chromatography prior to MS analysis. Six independent fractions obtained by MCX were then analyzed by nanoLC-MS (as described above). Data show a significant enrichment not only at the level of the total number of identified proteins (2297, Figure 1), but also confirmed that the pre-fractionation step was critical to access proteins with more TMDs. More than 350 proteins were assigned as plasma membrane proteins, including 92 receptors and proteins with numerous TMDs. Cell surface markers as myoferlin, endoglin and receptors (e.g. EGFR and Frizzled-6) were identified (Supporting Information Table ST3).

This proteomics platform has shown to be efficient in achieving a more complete description of the receptors present in hCSC membrane. Furthermore, the assessment of hCSC receptome was done using an Orbitrap Elite instrument (Thermo Scientific), its higher resolving power and sensitivity, together with a long 6 hours nanoLC run, improved and simplified sample analysis. Three replicates of enriched plasma membrane fraction, from three independent hCSC cultures (RI, RII and RIII) were analyzed. As before, proteins were identified with Proteome Discoverer 1.3, but a tolerance of 100 ppm or 1.2 Da was set for full MS or MS/MS spectra searches, respectively. A subsequent filter of 10 ppm tolerance for full MS was used, and peptide identifications validated using the probability ratio and the refined FDR method. Because numerous proteins were identified with one peptide only, an additional filter of FDR < 0.01 was applied to these proteins. A total of 3483 unique proteins were identified in the sum of the three runs, and among those 1242 were consistently identified in the three replicates (Figure 2; Supporting Information Table ST4), representing a common core of proteins which was consistently identified (Figure 2B). Nevertheless, a large number of proteins were identified in two (986 proteins) or in only one (1255 proteins) of the replicates. A substantial proportion (1318) of the annotated proteins contained one or more predicted TMDs, and from these 531 were identified in all triplicates, corresponding to approximately 40% of proteins identified in each list (Figure 3; Supporting Information Table ST4). These results reflect a significant enrichment in transmembrane proteins, as only approximately 27% of the total human proteome represent alpha-helical transmembrane proteins (containing one or more putative TMDs)[20]. Figures 2 and 3 summarize the dataset of proteins identified, namely concerning the presence of TMDs, cellular location and receptor function. As expected, although enrichment in plasma membrane proteins was achieved, a large contamination with proteins from other subcellular membranes was also present; this is in agreement with previous literature [11, 12, 16]. Nevertheless, by applying this subproteomic analysis to hCSC enriched plasma membrane fraction, we were able to achieve a larger coverage of hCSC receptome characterization. Up to this point, the list of receptors (and corresponding interaction partners) known to be relevant for hCSC activation is only a small fraction of the whole pool of molecules involved in this intricate process. Many others have escaped detection by the low sensitivity and target directed technologies used until now [13, 21, 22]. As referred before, the relatively low abundance of membrane proteins/receptors and the increased difficulty to identify integral membrane proteins due

to their hydrophobic nature has resulted in their under-representation in large proteomic studies. We took advantage of the combination of improved sample preparation options, nanoLC performance and state-of-the-art MS instrumentation, to translate it into a substantial increase in receptors identification in our dataset. This work and the data generated will critically contribute to the elucidation of the players beyond the mechanisms of hCSC activation and differentiation. Moreover, we are currently applying this optimized methodology, for identification of receptors in the characterization of other human cell types (e.g. mesenchymal stem cells). The reported workflow may constitute a key tool in the characterization of different human stem cell types (and other target cells) contributing to the interpretation of important biological mechanisms still undefined.

**Acknowledgements**

Authors acknowledge FP7 EU project CARE-MI (HEALTH-2009_242038) and the Portuguese Foundation for Science and Technology (PTDC/BBB-BIO/1414) for financial support. PGA is a recipient of the FCT fellowship SFRH/BPD/86513/2012. MALDI-TOF/TOF analyses were performed at the Mass Spectrometry Unit (UniMS), ITQB/iBET, Oeiras, Portugal. The data deposition to the ProteomeXchange Consortium was supported by PRIDE Team, EBI.

**Disclosure of potential conflicts of interest**

The authors have declared no conflict of interest.

**References:**

[1] WHO, *Fact sheet N°310*, WHO 2011.

[2] Beltrami, A. P., Barlucchi, L., Torella, D., Baker, M.*, et al.*, Adult cardiac stem cells are multipotent and support myocardial regeneration. *Cell* 2003, *114*, 763-776.

[3] Torella, D., Ellison, G. M., Karakikes, I., Nadal-Ginard, B., Resident cardiac stem cells. *Cell Mol Life Sci* 2007, *64*, 661-673.

[4] Torella, D., Indolfi, C., Goldspink, D. F., Ellison, G. M., Cardiac stem cell-based myocardial regeneration: towards a translational approach. *Cardiovasc Hematol Agents Med Chem* 2008, *6*, 53-59.

[5] Ellison, G. M., Torella, D., Karakikes, I., Nadal-Ginard, B., Myocyte death and renewal: modern concepts of cardiac cellular homeostasis. *Nat Clin Pract Cardiovasc Med* 2007, *4 Suppl 1*, S52-59.

[6] Torella, D., Ellison, G. M., Karakikes, I., Nadal-Ginard, B., Growth-factor-mediated cardiac stem cell activation in myocardial regeneration. *Nat Clin Pract Cardiovasc Med* 2007, *4 Suppl 1*, S46-51.

[7] Anversa, P., Kajstura, J., Leri, A., Bolli, R., Life and death of cardiac stem cells: a paradigm shift in cardiac biology. *Circulation* 2006, *113*, 1451-1463.

[8] Ellison, G. M., Torella, D., Dellegrottaglie, S., Perez-Martinez, C.*, et al.*, Endogenous cardiac stem cell activation by insulin-like growth factor-1/hepatocyte growth factor intracoronary injection fosters survival and regeneration of the infarcted pig heart. *J Am Coll Cardiol* 2011, *58*, 977-986.

[9] Nadal-Ginard, B., Kajstura, J., Leri, A., Anversa, P., Myocyte death, growth, and regeneration in cardiac hypertrophy and failure. *Circ Res* 2003, *92*, 139-150.

[10] Intoh, A., Kurisaki, A., Yamanaka, Y., Hirano, H.*, et al.*, Proteomic analysis of membrane proteins expressed specifically in pluripotent murine embryonic stem cells. *Proteomics* 2009, *9*, 126-137.

[11] Simoes, T., Charro, N., Blonder, J., Faria, D.*, et al.*, Molecular profiling of the human nasal epithelium: A proteomics approach. *J Proteomics* 2011, *75*, 56-69.

[12] Dormeyer, W., van Hoof, D., Mummery, C. L., Krijgsveld, J., Heck, A. J., A practical guide for the identification of membrane and plasma membrane proteins in human embryonic stem cells and human embryonal carcinoma cells. *Proteomics* 2008, *8*, 4036-4053.

[13] Moscoso, I., Tejados, N., Barreiro, O., Sepulveda, P.*, et al.*, Podocalyxin-like protein 1 is a relevant marker for human c-kit cardiac stem cells. *J Tissue Eng Regen Med* 2013.

[14] Lauden, L., Boukouaci, W., Borlado, L. R., Lopez, I. P.*, et al.*, Allogenicity of human cardiac stem/progenitor cells orchestrated by programmed death ligand 1. *Circ Res* 2013, *112*, 451-464.

[15] Larsen, M. R., Graham, M. E., Robinson, P. J., Roepstorff, P., Improved detection of hydrophilic phosphopeptides using graphite powder microcolumns and mass spectrometry: evidence for in vivo doubly phosphorylated dynamin I and dynamin III. *Mol Cell Proteomics* 2004, *3*, 456-465.

[16] Gu, B., Zhang, J., Wang, W., Mo, L.*, et al.*, Global expression of cell surface proteins in embryonic stem cells. *PLoS One* 2010, *5*, e15795.

[17] Bonzon-Kulichenko, E., Perez-Hernandez, D., Nunez, E., Martinez-Acedo, P.*, et al.*, A robust method for quantitative high-throughput analysis of proteomes by 18O labeling. *Mol Cell Proteomics* 2011, *10*, M110 003335.

[18] Navarro, P., Vazquez, J., A refined method to calculate false discovery rates for peptide identification using decoy databases. *J Proteome Res* 2009, *8*, 1792-1796.

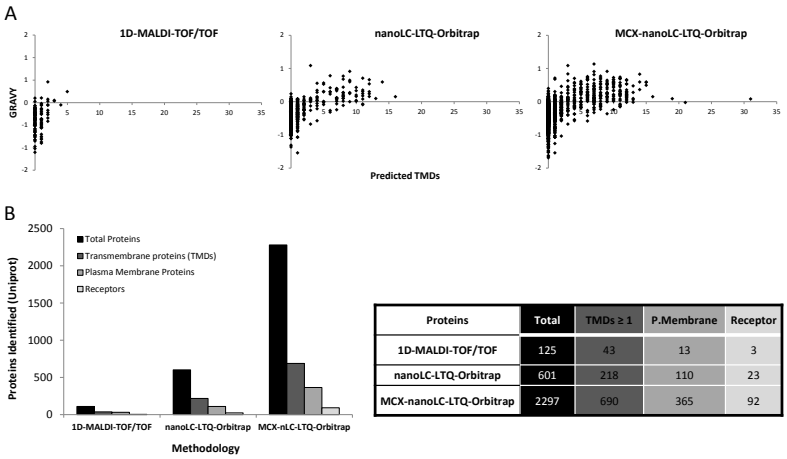
[19] Martinez-Bartolome, S., Navarro, P., Martin-Maroto, F., Lopez-Ferrer, D*., et al.*, Properties of average score distributions of SEQUEST: the probability ratio method. *Mol Cell Proteomics* 2008, *7*, 1135-1145.

[20] Almen, M. S., Nordstrom, K. J., Fredriksson, R., Schioth, H. B., Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biol* 2009, *7*, 50.

[21] Scobioala, S., Klocke, R., Kuhlmann, M., Tian, W*., et al.*, Up-regulation of nestin in the infarcted myocardium potentially indicates differentiation of resident cardiac stem cells into various lineages including cardiomyocytes. *FASEB J* 2008, *22*, 1021-1031.

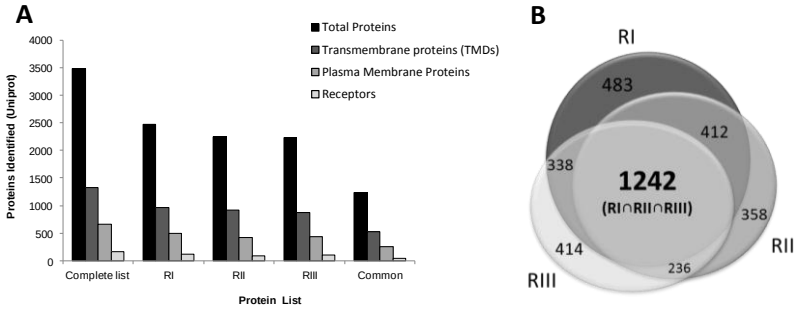
[22] D'Amario, D., Cabral-Da-Silva, M. C., Zheng, H., Fiorini, C*., et al.*, Insulin-like growth factor-1 receptor identifies a pool of human cardiac stem cells with superior therapeutic potential for myocardial regeneration. *Circ Res* 2011, *108*, 1467-1481.

Figure 1 – Optimization of hCSC receptome analysis. A- Number of predicted TMDs and GRAVY values of the identified proteins using three different approaches: 1D-MALDI-TOF/TOF, nanoLC-LTQ-Orbitrap and MCX-nanoLC-LTQ-Orbitrap. B- MCX-nanoLC-LTQ-Orbitrap approach yielded a higher number of total proteins identified, including higher number of proteins with predicted TMDs and receptors.




| Proteins | Total | TMDs ≥ 1 | P.Membrane | Receptor |
|---|---|---|---|---|
| 1D-MALDI-TOF/TOF | 125 | 43 | 13 | 3 |
| nanoLC-LTQ-Orbitrap | 601 | 218 | 110 | 23 |
| MCX-nanoLC-LTQ-Orbitrap | 2297 | 690 | 365 | 92 |

Figure 2 – hCSC receptome analysis using nanoLC-Orbitrap Elite. (A) Graphical representation of the number of proteins identified in hCSC enriched membrane fraction, *Classification according to IPA software; (B) Venn diagram representing the number of proteins identified in the three replicates (**http://bioinfogp.cnb.csic.es/tools/venny/**); (C) Classification and representativity of proteins identified in hCSC enriched membrane fraction.




C

| | Number of identified proteins | | | |
|---|---|---|---|---|
| | **Total** | **TMDs** | **Plasma Membrane*** | **Receptors*** |
| **Total list** | 3483 | 1318 (37.8%) | 663 (19.0%) | 167 (4.8%) |
| **RI** | 2475 | 970 (39.2%) | 491 (19.8%) | 123 (5.0%) |
| **RII** | 2248 | 914 (40.7%) | 416 (18.5%) | 93 (4.1%) |
| **RIII** | 2230 | 870 (39.0%) | 441 (19.8%) | 104 (4.7%) |
| **Common** | 1242 | 531 (42.6%) | 250 (20.1%) | 50 (4.0%) |

Figure 3 – (A) Subcellular location of the identified proteins in the three replicates according to IPA software database (NP- not present). (B) Number of predicted TMDs vs GRAVY values of the identified proteins. The methodology used allowed the identification of 1318 (531 common to all replicates) integral membrane proteins and also highly hydrophobic ones (645 proteins with GRAVY > 0; 228 common to all replicates).