Heart failure is one of the leading causes of death worldwide [1–4]. Current therapeutic strategies are inefficient and cannot cure this chronic and debilitating condition [5]. Ultimately, heart transplants are required for patient survival, but donor organs are scarce in availability and only prolong the life-span of patients for a limited time. Fibrosis is one of the main pathological features of heart failure [6,7], caused by inappropriate stimulation of fibroblasts and excessive extracellular matrix production. Therefore, an in-depth understanding of the cardiac fibroblast is essential to underpin effective therapeutic treatments for heart failure [5]. Fibroblasts in general have been an underappreciated cell type, regarded as relatively inert and providing only basic functionality; they are usually referred to as the “biological glue” of all tissues in the body. However, more recent literature suggests that they actively participate in organ homeostasis and disease [7,8].

We have recently uncovered a unique molecular identity for fibroblasts isolated from the heart [9], expressing a set of cardiogenic transcription factors that have been previously associated with cardiomyocyte ontogenesis. This signature suggests that cardiac fibroblasts may be ideal for use in stem cell replacement therapies, as they may retain the memory of where they derive from embryologically. Our data also revealed that about 90% of fibroblasts from both tail and heart origins share a cell surface signature that has previously been described for mesenchymal stem cells (MSCs), raising the possibility that fibroblasts and MSCs may in fact be the same cell type. Thus, our findings carry profound implications for the field of regenerative medicine. Here, we describe detailed methodology and quality controls related to the gene expression profiling of cardiac fibroblasts, deposited at the Gene Expression Omnibus (GEO) under the accession number GSE50531. We also provide the R code to easily reproduce the data quantification and analysis processes.

© 2014 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

Data in Brief

Microarray profiling to analyse adult cardiac fibroblast identity

Milena B. Furtado a,b,⁎, Hieu T. Nim a,b,⁎, Jodee A. Gould c, Mauro W. Costa a, Nadia A. Rosenthal a,b,⁎, Sarah E. Boyd a,b,⁎

a Australian Regenerative Medicine Institute, Monash University, VIC 3800, Australia
b Systems Biology Institute (SBI) Australia, Monash University, VIC 3800, Australia
c Monash Health and Translational Precinct (MHTP), Medical Genomics Facility (MGF), VIC 3168, Australia

Abstract

Heart failure is one of the leading causes of death worldwide [1–4]. Current therapeutic strategies are inefficient and cannot cure this chronic and debilitating condition [5]. Ultimately, heart transplants are required for patient survival, but donor organs are scarce in availability and only prolong the life-span of patients for a limited time. Fibrosis is one of the main pathological features of heart failure [6,7], caused by inappropriate stimulation of fibroblasts and excessive extracellular matrix production. Therefore, an in-depth understanding of the cardiac fibroblast is essential to underpin effective therapeutic treatments for heart failure [5]. Fibroblasts in general have been an underappreciated cell type, regarded as relatively inert and providing only basic functionality; they are usually referred to as the "biological glue" of all tissues in the body. However, more recent literature suggests that they actively participate in organ homeostasis and disease [7,8].

We have recently uncovered a unique molecular identity for fibroblasts isolated from the heart [9], expressing a set of cardiogenic transcription factors that have been previously associated with cardiomyocyte ontogenesis. This signature suggests that cardiac fibroblasts may be ideal for use in stem cell replacement therapies, as they may retain the memory of where they derive from embryologically. Our data also revealed that about 90% of fibroblasts from both tail and heart origins share a cell surface signature that has previously been described for mesenchymal stem cells (MSCs), raising the possibility that fibroblasts and MSCs may in fact be the same cell type. Thus, our findings carry profound implications for the field of regenerative medicine. Here, we describe detailed methodology and quality controls related to the gene expression profiling of cardiac fibroblasts, deposited at the Gene Expression Omnibus (GEO) under the accession number GSE50531. We also provide the R code to easily reproduce the data quantification and analysis processes.

© 2014 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

Direct link to deposited data

Deposited data can be found here: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE50531.

Experimental design, materials and methods

Mouse usage

Eight-week old C57Bl/6j (http://jaxmice.jax.org/strain/000664.html) adult male mice were obtained from Monash Animal Services (MAS) facility, Monash University.

Fibroblast isolation and culture

Animals were humanely killed using CO2 asphyxiation and hearts were perfused with 20 ml of Hank’s Balanced Salt Solution without calcium or magnesium (HBSS-Gibco) using a 20 ml syringe and 26 gauge needle.
needles. For perfusion, the needle was inserted into the left ventricular chamber, while the right atrial appendage was cut open to allow blood/buffer exit through the pulmonary circulation. This procedure removed excess blood from heart chambers. Prior to cell isolation, two wild-type mouse hearts per sample were finely minced using surgical scissors. Whole tails from the same animals were skinned and cut into 2–3 mm pieces. Both tissues were subjected to enzymatic digestion with 0.05% trypsin/EDTA (Gibco) at 37 °C under agitation for 40 min, washed in HBSS, cleared using 20 μm cell strainers for Falcon tubes (BD Biosciences) that retain undigested tissue, spun at 400 g for 5 min, resuspended in cell culture media (DMEM high glucose with glutamax (Gibco), supplemented with 10% foetal bovine serum (FBS-Gibco), 1× penicillin/streptomycin (Gibco) and 1× sodium pyruvate (Gibco)) and plated on 10 cm dishes (BD Biosciences). Next morning, plates were washed twice in phosphate buffered saline (PBS) without calcium or magnesium (Gibco) to remove debris. Media changes were performed every couple of days until day 5, after which cells were processed for RNA extraction. Cells were cultured in a humidified incubator with 5% CO2 at 37 °C.

RNA preparation

Cultured fibroblasts from heart and tail were processed for total RNA extraction using the mirVana kit (Ambion) as per manufacturer’s instructions. Briefly, 10 cm dishes containing heart or tail fibroblasts were washed twice in PBS (Gibco) and 500 μl of RNA lysis solution was added to dishes. Cell scrapers (BD Biosciences) were used to homogenise cell disruption, after which 1:10 (v/v) of miRNA homogenate additive was added. Samples were subjected to organic extraction using acid phenol:chloroform and 1.25 volume of ethanol was added to aqueous phases, which were passed through a filter cartridge, washed in kit wash solutions 1 and 2 and eluted in 50 μl of pre-heated (95 °C) elution solution. Following manufacturer’s recommendations for total RNA recovery, we have not enriched samples for small RNAs. All samples

![Fig. 1. Quality control (QC) analysis for samples used for microarray. A. Electropherogram shows typical plot for high quality total RNA, peaks for 18 s and 28 s ribosomal RNA subunits can be clearly seen for all samples and no intermediate peaks, equivalent to degradation products, are observed. B. Gel image showing expected migration pattern for intact RNA on tail and heart samples, highlighting 18 s (below 2000 bp band of ladder) and 28 s (below 4000 bp band of ladder) ribosomal RNA. Small peaks 25 and 200 nucleotides correspond to the loading marker and small RNAs, respectively. RNA Integrity Number (RIN) scores vary from a minimum of 0 (degraded) to a maximum of 10 (intact) total RNA. nt — nucleotide; FU — fluorescence intensity.](image-url)
were DNAse digested using DNAseI from the DNA-free kit (Ambion) for 20 min at 37 °C, after which 1:10 (v/v) DNAse inactivation reagent was added, mixed for 2 min at room temperature and spun down at maximum speed for 90 s in an Eppendorf centrifuge to pellet inactivation beads. Samples were transferred into clean tubes and further processed at the Medical Genomics Facility.

**Microarray study design**

Three replicate samples were profiled per condition (i.e. heart or tail fibroblasts).

**RNA labelling and hybridization**

Total RNA from 6 samples (3 tail and 3 heart replicas) was used for microarray analysis. All samples showed RNA integrity numbers (RINs) ranging between 9–10, as determined by the 2100 Bioanalyser (Agilent) (Fig. 1). 0.1 μg of total RNA was used to prepare Cyanine-3 (Cy3) labelled cRNA for hybridization. For labelling, the One-Color Low input Quick Amp labelling Kit (Agilent) was used according to the manufacturer’s instructions, after which labelled RNA was cleaned using by RNeasy column purification (Qiagen). Dye incorporation and cRNA yield were determined with the NanoDrop ND-1000 Spectrophotometer.

For chip hybridization, 600 ng of Cy3 labelled cRNAs were fragmented at 60 °C for 30 min in a reaction volume of 25 μl containing 1× Agilent Fragmentation buffer and 2× Agilent Gene Expression Blocking agent. Specific activities ranging between 16–18 pmol Cy3/μg cRNA were used for each sample. On completion, 25 μl of 2× Hi-RPM Gene Expression buffer (Agilent) was added. 40 μl of samples were hybridised on SurePrint G3 Mouse GE 8 × 60 K microarrays (Agilent) for 17 h at 65 °C in a rotating hybridisation oven (Agilent). Following hybridisation, microarrays were washed with GE wash buffer 1 (Agilent) for 1 min at room temperature and with GE wash buffer 2 at 37 °C for 1 min. Microarrays were scanned with an Agilent C, DNA microarray scanner at 3 μm resolution (scan area 61 × 21.6 mm), dye channel set to green, 20 bit tiff. The scanned images were analysed with Agilent Feature Extraction Software 11.0.1.1 to obtain background subtracted and spatially detrended Processed Signal intensities.

**Data normalisation and analysis**

Samples were processed by the Monash Health Translation Precinct (MHTP) Medical Genomics Facility, and run on Agilent SurePrint G3 mouse gene expression arrays (single colour), followed by analysis with GeneSpring 12.6, using quantile normalisation with no baseline transformation, logbase 2. Unpaired T test (p < 0.05, logFC > 2.0) with Benjamini–Hochberg correction was used to find differentially expressed genes.
expressed genes. These results have been reported in the original article [9].

For the purpose of improved accessibility, a similarity analysis was performed here using an open-source pipeline with R and Bioconductor. The raw single-channel signals were extracted with Agilent Feature Extraction Software 11.0.1.1. Non-uniform, saturated probes, and population outliers were filtered using the default “Compromised” option in GeneSpring GX12.6 (Agilent), with threshold raw signal of 1.0. At the end of this process, 6 text files (.txt) were exported for the data normalisation stage. This analysis has been deposited in GEO under the accession number (to be allocated by GEO) and is linked to the original samples GSM1220786-91 under accession number GSE50531.

Data normalisation was performed with R (http://www.r-project.org) using the publicly available Bioconductor packages (bioconductor.org) [10]. Three pre-processing and normalisation steps were performed: (1) use read.maimage function to extract the processedSignal values from the GeneSpring exported data files (Fig. 2A); (2) use averageList function to average the duplicate spots and log-2 transformation (Fig. 2B); and (3) use normalizeBetweenArrays function to perform quantile normalisation on all arrays (Fig. 2C).

Differential analysis between cardiac fibroblast and tail fibroblast samples was performed using the Bioconductor limma package [11], which applies linear models and differential expression functions to the transcriptomic data. With 6 normalised arrays having identical distributions (Fig. 2C), the limFit function identifies the genes that have differential expression between 3 cardiac fibroblast samples and 3 tail fibroblast samples. At a p-value threshold of 0.05, we identified a pool of 3924 differentially expressed entities (Fig. 2D). These entities were used for fold-change calculation (Fig. 2E), revealing 94 strongly up-regulated entities (16×) and 56 strongly down-regulated entities (−16×). Tables 1.1 and 1.2 detailed the gene symbols corresponding to the strongly up- and down-regulated entities, which revealed many cardiogenic genes, including Tcf21, Tbx20 and Gata4. The R code to reproduce this analysis is available in GEO.

**Discussion**

Our goal for the current experiment was to identify cardiac fibroblast-specific genes [1-8]. For this purpose, isolated cells were cultured for 5 days in order to remove carry-over of debris and DNA/RNA from dead cells.
cardiomyocytes present in fresh preparations. Contamination with cardiomyocyte structural markers has been reported in previous analyses [12], as the isolation method for compact tissue requires harsh cardiomyocytes present in fresh preparations. Contamination with cardiomyocytes would clarify this issue.

Upon visualisation of the raw data (Fig. 2A), we noticed that sample tail2 showed elevated overall signal intensity than others. As all samples were processed simultaneously and showed similar RIN scores (Fig. 1), we believe this finding is a technical problem inherent to the hybridisation step or chip composition. Nevertheless, this discrepancy did not impair our analysis.

As with many microarray datasets, data noise is an important issue. We performed differential analysis using the robust lmFit method, which is widely regarded as noise-tolerant in the bioinformatics community. The high-confidence entities evaluated by log fold-change and p-value (such as the cardiogenic genes Tcf21 and Tbx20) were subsequently validated using qPCR validation to confirm their biological relevance, as described by our main research article [9].

The microarray analysis pipeline described here made use of both the proprietary GeneSpring GX12.6 software and open source R packages. In addition to our original analysis [9], we have provided a second open-source analysis to facilitate the reproducibility of our study. Due to differences between methods (algorithms) provided by limma and GeneSpring, and the closed-nature of the Genespring software, there are discrepancies between the final fold-change calculations between the two analyses. However, although the exact values differed for entities found up- and down-regulated in the open source analysis, our genes of interest were similarly differentially regulated in both analyses. A range of bona fide haematopoietic genes were up-regulated in heart samples, for example Ccl19, Cdh21, and Cxcr4, while other haematopoietic markers, such as CD45, CD31 and CD34 were not significantly altered. Most of these genes have not been correlated with fibroblast biology yet, although expression of Cxcr4 has been previously reported in cardiac fibroblasts [13]. We have also not detected significant levels of CD31, CD45 or CD45 protein in our samples [9], supporting the argument that if present, haematopoietic cells are not a major contaminant in our cultures. However, leukocyte contamination in our cultures cannot be ruled out at this point. As our laboratory has previously characterised a population of resident macrophages in the heart [12], follow-up experiments using sorted CD90+ / CD45- / CD31- cells would clarify this issue.

Conflict of interest

The authors have no conflict of interest.

Funding

The ARC Discovery Grant DP130104792 was given to SEB and HK; and The ARC Stem Cell Grant and NHMRC Australia Fellowship were given to NR. The Australian Regenerative Medicine Institute is supported by grants from the State Government of Victoria and the Australian Government.

Acknowledgements

We acknowledge the use of facilities Monash Animal Services (MAS) and MHTP Medical Genetics Facility.

Appendix A Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gdata.2014.10.006.

References

[1] A.S. Go, D. Mozaffarian, E.J. Benjamin, J.D. Berry, M.J. Blaha, S. Dai, E.S. Ford, C.S. Fox, S. Franco, H.J. Fullerton, C. Gillespie, S.M. Hailpern, J.A. Heit, V.J. Howard, M.D. Huffman, S.E. Judd, B.M. Kissela, S.J. Kittner, D.T. Lackland, J.H. Lichtman, L.D. Lisabeth, R.H. Mackey, D.J. Magid, G.M. Marcus, A. Marelli, D.B. Matchar, D.K. McGuire, E.R. Mohler III, C.S. Moy, M. Mussolino, R.W. Neumar, G. Nichol, D.K. Pandey, N.P. Paynter, M.J. Reeves, P.D. Sorlie, J. Stein, A. Towfighi, T.N. Turan, S.S. Virani, N.D. Wong, D. Woo, M.B. Turner, American Heart Association Statistics Committee, Stroke Statistics Committee, American Heart Association. Circulation 129 (2014) 399–410.

[2] A.S. Go, D. Mozaffarian, E.J. Benjamin, J.D. Berry, M.J. Blaha, S. Dai, E.S. Ford, C.S. Fox, S. Franco, H.J. Fullerton, C. Gillespie, S.M. Hailpern, J.A. Heit, V.J. Howard, M.D. Huffman, S.E. Judd, B.M. Kissela, S.J. Ritter, D.T. Lackland, J.H. Lichtman, L.D. Lisabeth, R.H. Mackey, D.J. Magid, G.M. Marcus, A. Marelli, D.B.
A systematic analysis for the global burden of disease study 2010. Lancet 380 (2012) 2224–2260.

C.D. Mathers, D. Loncar, Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med. 3 (2006) e442.

E.B. Schelbert, G.C. Fonarow, R.O. Bonow, J. Butler, M. Gheorghiade, Therapeutic targets in heart failure: refocusing on the myocardial interstitium. J. Am. Coll. Cardiol. 63 (2014) 2188–2198.

K.B. Schuetze, T.A. McKinsey, C.S. Long, Targeting cardiac fibroblasts to treat fibrosis of the heart: focus on hdcas. J. Mol. Cell. Cardiol. 70 (2014) 100–107.

S. Van Lint, K. Miteva, C. Tschope, Crosstalk between fibroblasts and inflammatory cells. Cardiovasc. Res. 102 (2014) 258–269.

T. Moore-Morris, N. Guimarães-Camboa, I. Banerjee, A.C. Zambrin, T. Kisseleva, A. McAlpine, A.M. Costa, E.A. Pinto, S.S. Virani, N.D. Wong, D. Woo, M.B. Turner, American Heart Association Statistics C, Stroke Statistics S. Heart disease and Stroke Statistics-2014 update: a report from the American Heart Association. Circulation 129 (2014) e28–e292.

S.S. Lim, T. Von, A.D. Flaxman, G. Danaei, K. Shibuya, H. Adair-Rohani, M. Ahamm, H.R. Anderson, K.C. Andrews, M. Aryee, C. Atkinson, L.J. Bacchus, A.N. Bahalim, K. Balakrishnan, J. Balmes, S. Barker-Collo, A. Baxter, M.L. Bell, J.D. Blore, F. Blyth, C. Bonner, G. Borges, R. Bourne, M. Boussinesq, M. Brailer, P. Brooks, N.G. Bruce, B. Brunekreef, C. Bryan-Hancock, C. Burnhilder, F. Bull, R.T. Burnett, J.J. Byers, A. Calabria, J. Carapetis, E. Carsebro, Z. Chafe, S. Charlson, H. Chen, J.S. Chen, A.T. Cheng, J.C. Child, A. Cohen, K.E. Colson, R.C. Cowie, S. Darby, S. Darling, A. Davis, L. Degenhardt, F. Dentener, D.C. Del Rio, J. Devries, K. Dherani, C. Diorio, J.P. Djuric, J. Dorigo, J. Driscoll, K. Edmond, S.E. Eissenberg, B.J. Ellerin, S.J. Eben, A.C. Eiser, S. Fahimi, G. Farzadfar, S. Ferrari, M.J. Finucane, S. Flaxman, F.G. Fowkes, G. Freedman, M.K. Freeman, E. Gakidou, S. Ghosh, E. Giovannucci, G. Gmel, K. Graham, R. Grainger, B. Grant, D. Gunnell, H.R. Gutierrez, W. Hall, H.W. Hoek, A. Hogan, H.D. Hsogood III, D. Hoy, H. Hu, B.J. Hubbell, S.J. Hutchings, S.E. Ibeanusi, G.L. Jacklyn, R. Jaspers, J.B. Jonas, H. Kao, J.A. Kanis, N. Kassebaum, N. Kawakami, Y.H. Khiang, S. Khatibzadeh, J.P. Khoo, C. Kok, F. Laden, R. Lalloo, Q. Lan, T. Lathlean, J.L. Leasher, J. Liece, Y. L, J.K. Lin, S. Lipshutz, S. London, R. Lzano, Y. Lu, J. Mak, R. Malekzadeh, L. Mallinger, W. Marques, L. March, R. Marks, R. Martin, P. McGale, J. McGrath, S. Mehta, C.A. Mensah, T.K. Merriman, R. Michal, C. Michaud, Y. Mishra, K. Mohd Hanafiah, A.A. Mokdad, L. Morawska, D. Mozaffarian, T. Murphy, M. Naghavi, B. Neal, P.K. Nelson, J.M. Nolla, R. Norman, C. Olives, S.B. Olmer, L. Orchard, R. Osmoh, B. Oto, R. Page, K.D. Pandey, C.D. Parry, E. Passmore, J. Patra, N. Pearce, P.M. Pelizzari, M. Petrzed, M.R. Phillips, D. Pope, C.A. Pope III, J. Powles, M. Raz, H. Razavi, B. Ostro, A. Page, K.D. Pandey, C.D. Parry, E. Passmore, J. Patra, N. Pearce, P.M. Pelizzari, M. Petrzed, M.R. Phillips, D. Pope, C.A. Pope III, J. Powles, M. Raz, H. Razavi, B. Ostro, A. Page, K.D. Pandey, C.D. Parry, E. Passmore, J. Patra, N. Pearce, P.M. Pelizzari, M. Petrzed, M.R. Phillips, D. Pope, C.A. Pope III, J. Powles, M. Raz, H. Razavi, B. Ostro.