Genome Biology

RESEARCH Open Access

Pro-regenerative fingerprints identified in a sub-population of adult mouse cardiomyocytes by integrative single-cell proteomics

Consuelo Marín-Vicente[1,2*], Cristina Villa del Campo[2], Enrique Calvo[3,4], Jose Manuel Rodríguez[3,4], Rocío Sierra[2], Sandra Martín-Salamanca[2], Carlos Torroja[5], Akos Végvári[6], Roman A. Zubarev[6,7], Miguel Torres[2,4*] and Jesús Vázquez[1,3,4*]

*Correspondence: cmarin@cnic.es; mtorres@cnic.es; jesus.vazquez@cnic.es

[1] Cardiovascular Proteomics Group, Centro Nacional de Investigaciones Cardiovasculares (CNIC), Madrid, Spain
[2] Genetic Control of Development and Organ Regeneration Group, Centro Nacional de Investigaciones Cardiovasculares (CNIC), Madrid, Spain
[3] Proteomics Unit, Centro Nacional de Investigaciones Cardiovasculares (CNIC), Madrid, Spain
[4] CIBER de Enfermedades Cardiovasculares (CIBERCV), Madrid, Spain
[5] Bioinformatics Unit, Centro Nacional de Investigaciones Cardiovasculares (CNIC), Madrid, Spain
[6] Division of Chemistry I, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden
[7] Department of Pharmacological & Technological Chemistry, I.M. Sechenov First Moscow State Medical University, Moscow 119146, Russia

Abstract

Background: Recent advances in instrument sensitivity and sample preparation techniques are significantly improving the ability to study the heterogeneity of cell populations at the single-cell level by mass spectrometry-based proteomics. Integrating multiple layers of cellular data offers additional and yet unexplored insights in single-cell proteomics. In regenerative research, cardiomyocyte proliferation driven by the overexpression of the Myc transcription factor has been described, however, it has not yet been investigated at single-cell resolution.

Results: By using an optimized adult cardiomyocytes isolation procedure from mouse models and taking advantage of the integrative capabilities of the iSanXoT application, we are able to minimize batch effects and cell size-related biases, obtain quantitative subcellular compartment information and detect protein alterations within subcellular compartments, as it had not yet been defined for this methodology. This approach enhances data quantification accuracy and facilitates biological interpretation. We show that the Myc transcription factor switches the expression profile of metabolic enzymes and expands a subpopulation of adult cardiomyocytes with a pro-regenerative signature.

Conclusions: We demonstrate that different layers of information can properly pattern the proteomic phenotype of single-cells. The analysis of single-cardiomyocyte data with the integrative statistical framework of iSanXoT provided important clues to understand the impact of Myc transcription factor in provoking different immaturity and pro-regenerative signatures in adult mouse cardiomyocytes. The cellular heterogeneity exerted by mouse cardiomyocytes upon Myc overexpression demonstrates the relevance of conducting regenerative studies at the single-cell level for precisely defining the amplitude of this response in the heart.

Keywords: Mass- spectrometry, Proteomics, Single-cell, Subcellular compartment, Heart, Regeneration, Myc transcription factor

Background

High-throughput single-cell strategies have become a powerful tool for identifying cellular populations in various biological contexts. Unlike bulk analyses, which provide an average signal that may mask individual cell diversity, single-cell data enable the detection of heterogeneity within the same kind of cells. Single-cell transcriptomics has been the most commonly used approach for characterizing individual cells and mapping distinct populations based on RNA expression profiles [1–3].

Mass spectrometry-based single-cell proteomics is emerging as an alternative tool to discern subpopulations of cells. Progress in instrument sensitivity and in sample preparation approaches are enhancing our ability to study cellular heterogeneity at the protein level with increasing throughput and depth [4–9]. Single-cell proteomics information can reveal functional cell clusters that may not correlate with those defined by the mRNA profile [6, 10, 11], providing complementary information about cellular heterogeneity.

To correctly interpret single-cell proteomics data various stages of the experimental workflow are relevant, including cell isolation, sample preparation, data acquisition, and analysis [12]. The ScopeMS and Scope2 sample preparation methods offer increased sensitivity and coverage taking advantage of multiplexing by introducing an isobaric labelled "carrier proteome" [4, 5, 10]. However, multiplexing inherently introduces batch effects that need to be removed before cell clustering. Data interpretation can also be biased by cellular characteristics, such as cellular size. Typically, transcript expression or protein abundance values normalized to each cell or to total experimental count are used for cellular clustering analyses [4, 10, 13–16]. However, this procedure does not necessarily remove the batch effect or discern whether the observed changes arise from diversity in cell size or represent true abundance differences between the cells [17]. In the field of single-cell proteomics, choosing the adequate strategy to normalize the protein abundance data is essential to elucidate the biological origin of the observed differences. This problem has not been adequately addressed yet in literature.

In this work, we evaluate the performance of iSanXoT [18] to analyze single-mouse cardiomyocyte data. iSanXoT is a user-friendly and fully customizable application based on the SanXoT package [19], that enables sequential processing of quantitative mass spectrometry data at any user-defined level using the Generic Integration Algorithm (GIA). At each integration step, the GIA models the variance structure of quantitative variables and applies standardization, including normalization, by transforming them into Z-values. The Systems Biology Triangle (SBT) algorithm [20] included within iSanXoT also opens the possibility to incorporate additional tailored layers of information, such as subcellular compartmentalization of proteins. This is a novel concept in data handling that to our knowledge has not been applied in the field of single-cell proteomics.

We applied iSanXoT to study cellular heterogeneity in single adult cardiomyocytes with or without overexpression of the transcription factor Myc. In mammals, the regenerative capacity of the heart is restricted to a very narrow post-natal window. Cardiac damage suffered after the regenerative window is closed results in irreversible loss of cardiomyocytes, due to their inability to divide [21]. Recent studies show that overexpression of the transcription factor Myc promotes cardiomyocyte proliferation, potentially

enhancing regenerative capacity [22]. However, the effects of Myc overexpression are not necessarily homogeneous across the cardiomyocyte population and whether pro-regenerative sub-populations of cardiomyocytes exist in the adult mammalian heart remains unknown. Myc overexpression has been linked to several cellular phenomena, including a metabolic switch that may support a pro-regenerative phenotype [23] and a cardiomyocyte size reduction [24, 25]. We leverage this well-established effect to validate our analytical approach, testing it at the single-cell level in the context of the cellular, molecular and metabolic changes induced by Myc overexpression in cardiomyocytes. Only a limited number of studies have applied single-cell proteomics to cells isolated from tissue, while the majority of single-cell proteomics studies have focused on cells isolated from cultured samples [16, 26–29]. This study is among the first of its kind in the field [16, 26], and the first to apply single-cell proteomics to a large population of genetically modified mammalian cardiomyocytes directly isolated from the heart, representing a novel approach in regenerative cardiovascular research.

Results

Analysis of single-cell proteomics data using iSanXoT mitigates batch effects and compensates for cell size variations

Currently, addressing batch effects and cell size effects remains an open challenge in single-cell proteomics [12, 17, 30, 31]. To evaluate the effectiveness of iSanXoT in handling these factors, we generated proteomics data on control and Myc-overexpressing adult cardiomyocytes, obtained by using a previously reported genetic mouse model [24, 32]. We first performed a pilot experiment comprising 48 control and 48 Myc-overexpressing cardiomyocytes (dataset 1, [33]). To ensure cell viability, these cells were manually isolated (Fig. 1a and "Methods" section) and subjected to protein extraction and digestion. The resulting peptides were labelled with 10-plex TMT (Fig. 1b), following two experimental designs for tag randomization (Fig. 1c).

Analysis of raw reporter ion intensities revealed that TMT-channels containing control cardiomyocytes consistently showed higher intensities than those containing Myc-overexpressing cardiomyocytes (Fig. 1d), indicating higher protein amounts in control cardiomyocytes. This observation is also evident when comparing median intensity values across control and Myc-overexpressing cardiomyocytes (Fig. 2a). These results are in good agreement with the differences in cell size observed by optical microscopy (Fig. 1e), where control cardiomyocytes appear consistently larger than Myc-overexpressing ones, as previously described [24, 25]. Therefore, these results suggest that the raw proteomics data reflect cell size differences between the two cell groups.

Besides the cell size-related bias, unsupervised Principal Component Analysis (PCA) using the raw reporter ion intensities also revealed a strong TMT batch effect influencing cell distribution (Fig. 2b and c and Additional file 1: Fig. S1a showing the rest of the principal components) that required a proper normalization. This TMT batch effect reflects the compilation of cells in different groups of TMTs, where each of them contains 8 different cells labelled with different reporter channels that are procesed and analyzed together in the same LC–MS run. TMT channel was not driving the clustering of cells as it is seen in Additional file 1: Fig. S2. Cell size and TMT batch effects explained around 33% and 50% of data variance, respectively, according to Limma analysis (Fig. 2g

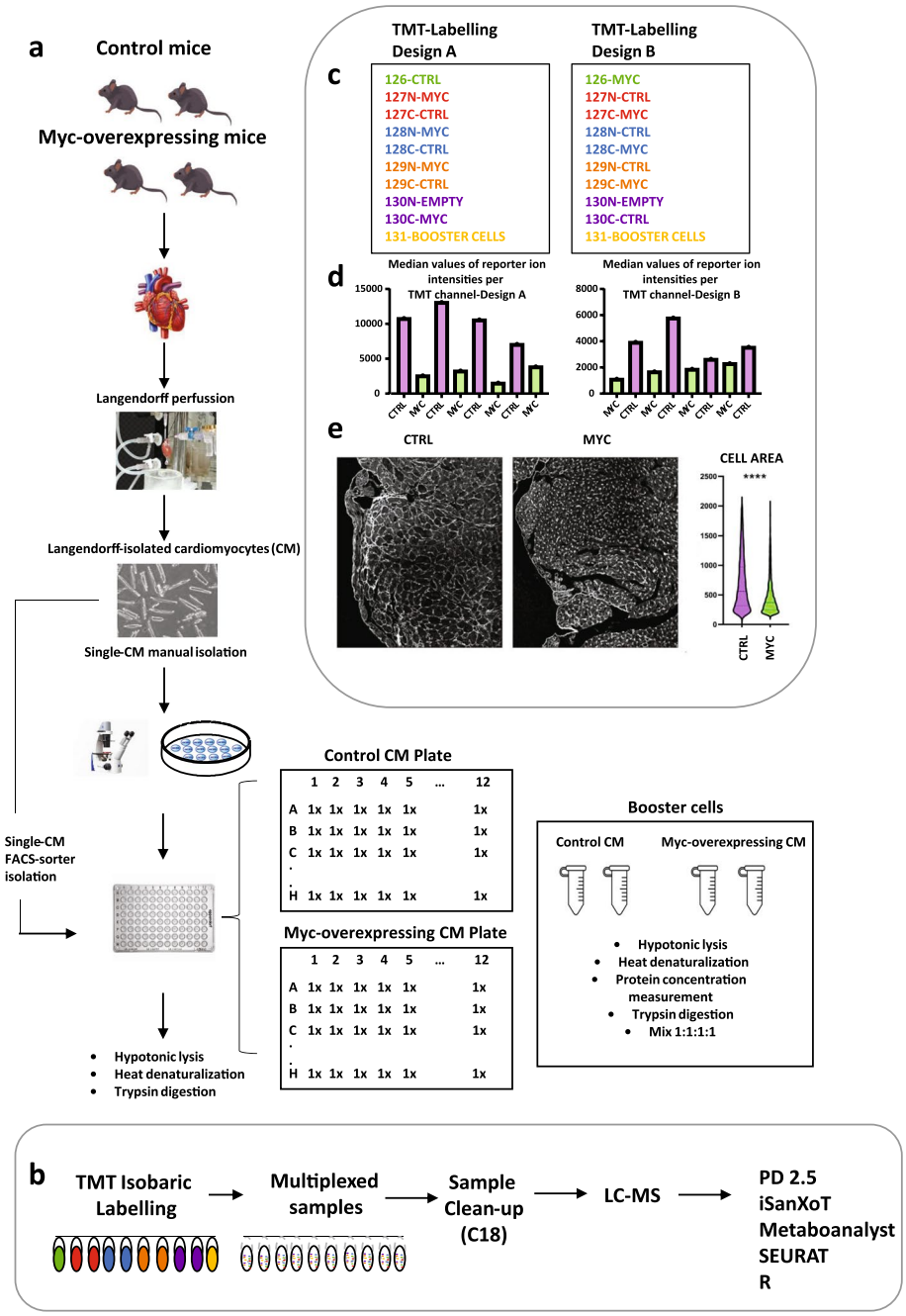

Fig. 1 General overview of our single-cardiomyocyte proteomics workflow. **a** Scheme of the methodological procedure followed for single-cardiomyocyte isolation and sample digestion. **b** Diagram of the steps for TMT labelling and data analysis of samples after digestion. TMT-isobaric labelled samples are multiplexed prior sample desalting and data acquisition by LC–MS. **c** Design A and B for TMT labelling of single-cardiomyocytes and booster channel. In the pilot analysis, 12 different 10-plex TMTs were used. **d** Median values of reporter ion intensities for single-cardiomyocytes in two experimental examples of Design A and Design B. Each bar represents the median values of reporter ion intensities of one cell. **e** Optical images of tissue sections from control and Myc-overexpressing hearts, along with quantification of cardiomyocyte cell area [25]

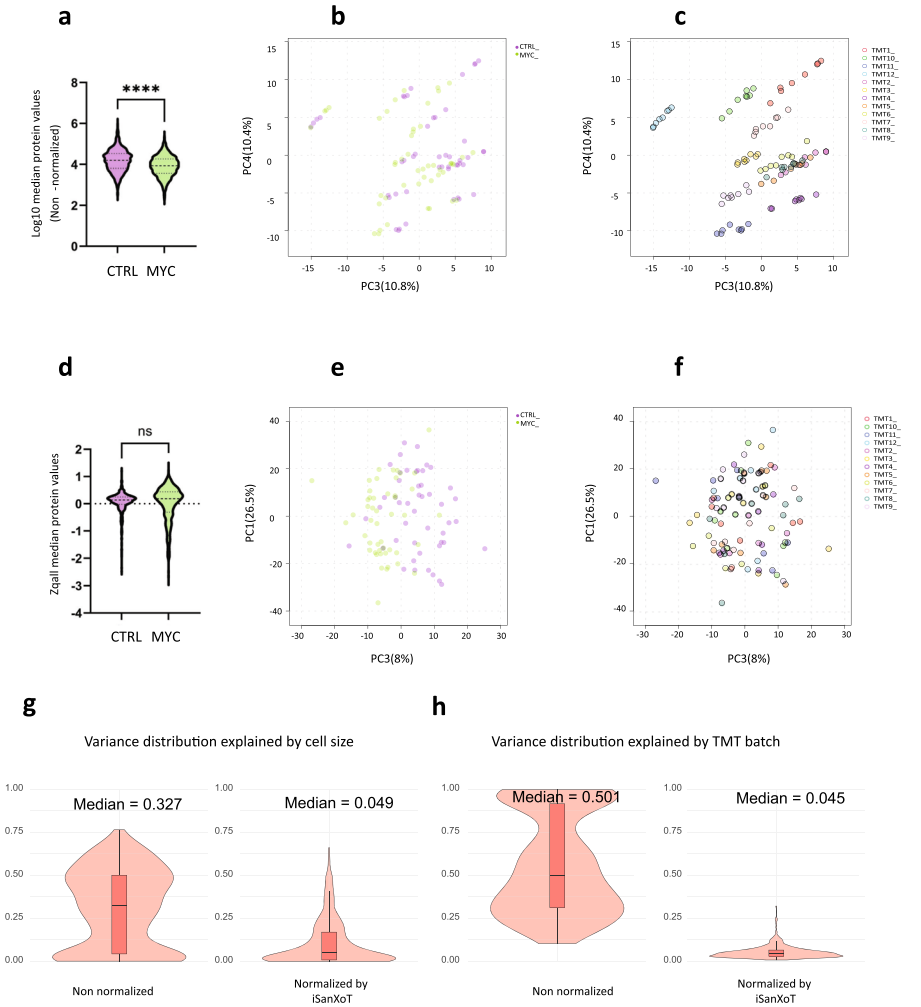

Fig. 2 Normalizing strategies for analyzing mass spectrometry-based single-cell proteomics data. **a** Violin plots showing the log10 transformed median values of proteins in control and Myc-overexpressing cardiomyocytes with non-normalized data. **b** PCA distribution of cells based on non-normalized protein expression values highlighting the cell type as control and Myc-overexpressing cardiomyocytes. **c** PCA distribution of cells based on non-normalized protein expression values highlighting the TMT where each cell belongs to. **d** Violin plot showing the Zqall median values, where Zqall represents the relative abundance changes of proteins, comparing control and Myc-overexpressing cardiomyocytes. **e** PCA representing the separation of control and Myc-overexpressing cardiomyocytes based on Zqall values. **f** PCA representing the separation of cardiomyocytes based on Zqall values and indicating the TMT in which each cardiomyocyte was analyzed. In every section, violin representations include the median values of 503 proteins out of 48 control cardiomyocytes and 48 Myc-overexpressing cardiomyocytes. PCA representations include the 48 control cardiomyocytes and 48 Myc-overexpressing cardiomyocytes used in the violin plots. Comparison between groups was performed by Student's t-test. (****: $p < 0.0001$). In this analysis, 12 different 10-plex TMTs were used. PERMANOVA analysis in the absence of normalization shows that F-value = 2.67 and p-value = 0.001. PERMANOVA analysis after applying iSanXoT shows that F-value = 4.05 and p-value = 0.001. **g** Limma analysis has been applied to analyze the influence of the cell size on the variance of protein values along samples when using non normalized data or data normalized by iSanXoT. **h** Limma analysis has been applied to analyze the influence of TMT batches on the variance of protein values along samples when data are non normalized or normalized by iSanXoT. For the cases where data are normalized by iSanXoT and explained variance is close to 0, p-value based on permutation test is computed to evaluate whether the explained proportion is significantly different from 0. In both cases, p-value > 0.05 (p-value = 0.961 (cell size), p-value = 1 (TMT)). See also Additional file 1: Figs. S1 and S3

and h). Current strategies in the field typically rely on normalizing data either by cellular protein abundance or by the total protein abundance across each experimental batch, a problem highlighted in several publications [12, 16, 30, 31, 34]. When normalizing by cellular protein abundance (i.e. by total reporter ion intensities within each cell), the difference between median protein abundances of control and Myc-overexpressing cardiomyocytes was corrected, and cell size explained only 2% of the median variance of the data (Additional file 1: Fig. S3a and S3g); however, the TMT batch effect found in PCA analysis remained (Additional file 1: Fig. S3b and S3c, S3h and Fig. S1c showing the rest of the principal components). The batch effect caused by TMT explained around the 65% of the variance found in the data (Additional file 1: Fig. S3h). To analyze whether the normalized data could differentiate the two cell populations according to their multivariate composition, we performed a PERMANOVA. This analysis indicated that separation between control and Myc-overexpressing cardiomyocytes was not significant ($F = 0.704$, p-value $= 0.765$). On the other hand, we tried to adjust for TMT batch effects by normalizing to the total protein abundance of all channels included within the same TMT batch. With the resulting data, unsupervised PCA showed separation between control and Myc-overexpressing cardiomyocytes in some components (Additional file 1: Fig. S3e and Fig. S1d showing the rest of the principal components), and PERMANOVA indicated that the separation between cell groups was significant ($F = 2.67$, p-value $= 0.001$). However, Limma analysis showed that around 40% and 45% of data variance was explained by cell size and TMT batches, respectively (Additional file 1: Fig. S3g and S3h), indicating that these effects were not corrected. The fact that the difference between the median protein abundance values of control and Myc-overexpressing cardiomyocytes remained significant (Additional file 1: Fig. S3d) suggested that cell separation in Additional file 1: Fig. S3e was primarily driven by cellular size.

In an effort to eliminate cell size and batch biases, we took advantage of the advanced capabilities of iSanXoT built-in GIA for data standardization [18] (see "Methods" section for further details about the steps followed by iSanXoT). Since this package has been never applied before to single-cell proteomics data, we first analyzed whether this statistical framework was adequate to model this kind of data. We first confirmed that the variance calibration module of iSanXoT [18] was able to adequately model the distribution of scan variances (Additional file 1: Fig. S4a). We then integrated quantitative data in each cell from scan to peptide, from peptide to protein and from protein (q) to grand mean (all), finding that in all cases the distribution of Z values had a good agreement with the expected standard distribution N(0, 1) (Additional file 1: Fig. S4b). Besides, the residual variances obtained at each of the integration steps were similar to those obtained when performing conventional proteomics analysis (see for instance [35]) (Additional file 1: Fig. S4b, violin plots). These data showed that iSanXoT was able to accurately model quantitative results from single-cell proteomics data.

When analyzing the median standardized Z values at the protein level (Zqall) no significant differences between control and Myc-overexpressing cardiomyocytes were found (Fig. 2d), while cell size and TMT batch effects were significantly reduced, explaining less than 5% of the data (Fig. 2e, f, g, h and Additional file 1: Fig. S1b). Besides, a permutation test showed that these variances were not significantly different to zero (p-value $= 0.961$ for cell size and p-value $= 1$ for TMT batch effects) (see captions in Fig. 2), suggesting

that the remaining contribution of these variances were negligible. Using Zqall values, we found a clear separation of both types of cardiomyocytes specially driven by principal component 3 (Fig. 2e and Additional file 1: Fig. S1b) and PERMANOVA indicated that the separation between the two cell groups was significant (F = 4.05, *p*value = 0.001). Overall, these results suggest that the variance standardization algorithm of the GIA is particularly efficient to normalize quantitative data for single-cell proteomics, being able to avoid both batch and cell size effects.

The method for isolating cardiomyocytes is critical in single-cell proteomics

Once we validated iSanXoT as an adequate quantitative workflow to process single-cell data using the pilot dataset [33], we analyzed its performance for the analysis of a larger dataset (analytical dataset, [36]). Due to the high number of cells that are required in single-cell analyses, automated isolation methods are preferable. Among these, cell isolation using conventional fluorescence-activated cell sorting (FACS) is the most popular method. However, adult cardiomyocytes are cell types of relatively large size, and their isolation using FACS has been reported to have limited efficiency and to provoke stress to the cardiomyocytes during the procedure [37, 38]. For this reason, we decided to compare the single-cell proteomics results obtained using either automated FACS-isolation or a manual isolation that ensures cardiomyocyte integrity prior to their lysis (Fig. 1a and "Methods" section for details). These two methods were tested on cell populations composed each one by 56 control and 56 Myc-overexpressing cardiomyocytes (datasets 2 for FACS isolation [39] and 3 for manual isolation [36]), see "Methods" section for details). Single-cell proteomes were obtained and processed with iSanXoT, as above. The results showed that control and Myc-overexpressing cardiomyocytes isolated by FACS were not separated by unsupervised PCA (Fig. 3a), and PERMANOVA indicated that the separation was not significant (*p*-value = 0.184). In contrast, the PCA analysis on manually-isolated cells was able to separate the two conditions (Fig. 3b) and PERMANOVA statistics were statistically significant (*p*-value = 0.001). Evaluation of data showed that average of number of peptides and proteins per cell, percentage of missing values and the number of proteins present in at least 70% of cells was similar in both

(See figure on next page.)

Fig. 3 Comparison of two strategies for isolating single-cardiomyocytes. **a** PCA representing the separation of control and Myc-overexpressing cardiomyocytes isolated by FACS-sorter. **b** PCA representing the separation of control and Myc-overexpressing cardiomyocytes manually isolated. 948 common proteins to both separation techniques have been compared considering their identification in at least 16 cells, that would include proteins present in at least 1 TMT per condition. Metaboanalyst filters out proteins with constant imputed values under standard deviation and mean filtering criteria. Each PCA representation includes 56 control and 56 Myc-overexpressing cardiomyocytes. A total of 14 different 10-plex TMTs were used in each of the analyses. The 10-plex TMTs used in the manual isolation were included in the final analytical dataset. PERMANOVA test results: In Fig. 3a, F-value = 1.72 and *p*-value = 0.184. In Fig. 3b, F-value = 31.741 and *p*-value = 0.001. **c** and **d** Violin plots showing the distribution of Zqall values of proteins related to fatty acid oxidation that show to be discriminant in the differential analysis when manual isolation of cardiomyocytes is conducted. **c** Represents the case when FACS isolation is applied and **d** when manual isolation is performed. **e** Violin plots showing the distribution of the differential reporter ion intensities when comparing control and Myc-overexpressing cardiomyocytes isolated with FACS or manually. In all cases, control cells are represented in violet and Myc-overexpressing ones in green. Comparison between control and Myc-overexpressing groups was performed by Student's t-test (*: $p < 0.05$; ***: $p < 0.001$; ****: $p < 0.0001$). **f** and **g** Volcano plots where discontinuous lines separate discriminant proteins when isolating cells by FACS (**c**) or manually (**d**) (thresholds: *p*value < 0.05 and fold change (FC) > 2)

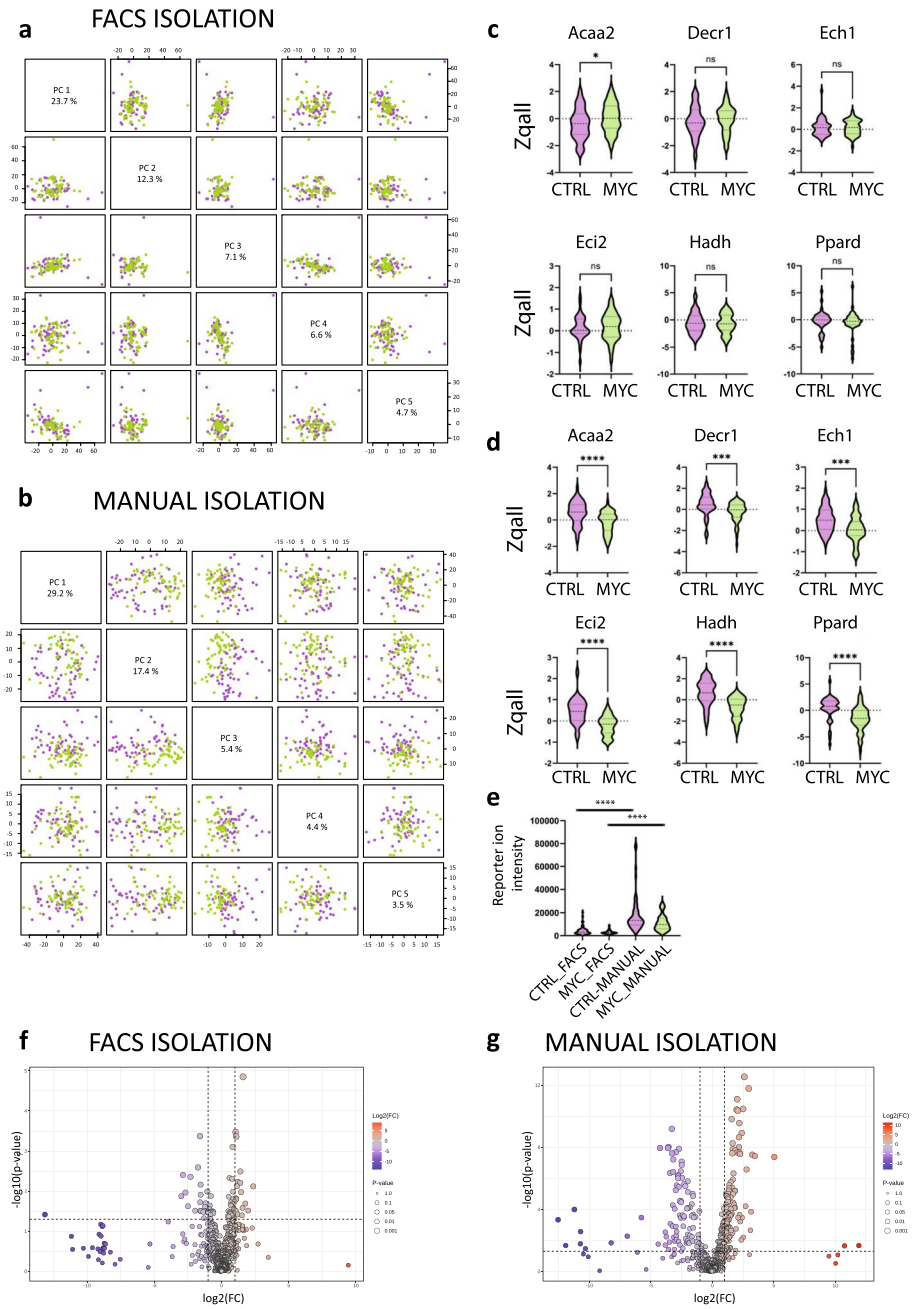

Fig. 3 (See legend on previous page.)

manual and FACS isolation methods (Additional file 1: Table S1). However, the number of proteins present in 100% of cells differed significantly, being 32 when isolating by FACS and 223 when isolating manually, suggesting that manual isolation was more reproducible (Additional file 1: Table S1). In addition, differential expression analysis showed that FACS-isolated cardiomyocytes can be discriminated with a lower number of proteins than when comparing control and Myc-overexpressing cardiomyocytes isolated manually (Fig. 3f and g). Finally, since beta oxidation of fatty acids is known to be the primary metabolic pathway that provides energy to adult cardiomyocytes and that is

reduced upon Myc overexpression [23], we took this already validated fact as a hallmark and inspected the behaviour of these proteins in the two isolation procedures. We found that representative fatty acid oxidation-related proteins showed low- or non-significant changes when comparing both types of cells isolated by FACS; however, all these proteins showed highly significant changes when comparing manually isolated cells (Fig. 3c and d). These results suggest that FACS isolation may introduce cellular variations that hinder the distinction between the two cell types during subsequent analysis.

One possible explanation is that FACS affects cell integrity. To evaluate this factor, we analyzed the overall distribution of non-normalized reporter ion intensities; we found that the values in FACS sorted cells were significantly smaller than those produced in manually isolated cells (Fig. 3e), suggesting a perturbation of the cell size that may interfere with the analysis. Overall, these comparisons evidence the considerable impact that the method of cell isolation has on the generation of biologically interpretable single-cell proteomics data.

Myc overexpression reduces fatty acid oxidation proteins in adult cardiomyocytes

After establishing a procedure for cardiomyocyte isolation and data processing, we applied the overall workflow to an analytical set [36] composed by 323 control and 324 Myc-overexpressing manually isolated cardiomyocytes and initially analyzed them as a pseudobulk, that is, comparing the two experimental groups, one of control and one of Myc-overexpressing cardiomyocytes. We quantified a total of 1624 proteins in the single-cell analytical dataset [36], from which 754 proteins were quantified with two or more peptides and 483 proteins were quantified in at least 30% of cells. We quantified an average of 440 proteins per cell. The median Zqall values produced by iSanXoT showed no significant difference between control and Myc-overexpressing cardiomyocytes and Limma analysis showed a neglible impact of both cell size and TMT batch on data variance, demonstrating again that iSanXoT efficiently corrects for cell size and TMT bias (Fig. 4a and f). The PCA analysis detected a tendency to separate both cell types (Fig. 4b), and PERMANOVA analysis reported the group separation as significant ($F = 9.81$, p-value $= 0.001$). Differential expression analysis showed that fatty acid oxidation category emerges as a biological process when comparing control and Myc-overexpressing cardiomyocytes, validating the known effect of Myc on adult cardiomyocytes metabolism (see protein lists in Zenodo link [40]). The separation between the two complete cell groups was clearly observed by a supervised multivariate analysis such as Partial Least Squares Discriminant Analysis (PLS-DA) (Fig. 4c). This supervised strategy is commonly used for feature selection and classification to highlight the differences between cell groups since it leverages covariance structure across proteins, detecting coordinated changes that univariate tests may miss [41–43]. The separation obtained under the PLS-DA analysis suggests the presence of proteins that can distinguish between both predefined groups of cells. To control potential overfitting of the model, cross-validation analysis was performed, showing consistent R^2 and Q^2 values across components, with Q^2 exceeding 0.54 and closely tracking R^2. The Q^2 curve plateaued after two components, indicating stable predictive performance without overfitting. The classification accuracy increased smoothly from 0.69 (1 component) to 0.93 (5 components), supporting a well-generalized model (Additional file 1: Table S2). We extracted the proteins

that contributed to this separation based on their PLS-DA VIP scores (113 proteins, see correspondant file in Zenodo link [40]) and performed functional interaction network analysis to investigate their nature. Besides other typical processes known to be involved in cardiomyocyte biology, such as cytoskeleton organization or cardiac muscle contraction, we observed that fatty acid beta-oxidation was one of the categories that discriminate both types of cardiomyocytes (Fig. 4d and Additional file 1: Fig. S5a). Among the proteins belonging to this category, 10 of them were discriminant proteins when comparing control and Myc-overexpressing cardiomyocytes and were identified in 100% of the cells; these proteins were consistently found at statistically significant lower levels in Myc-overexpressing cardiomyocytes (Fig. 4e). This decrease in fatty acid beta-oxidation proteins reproduces at the basal level the observed metabolic pattern described in previous studies following Myc overexpression in mouse hearts [23]. Hence, the proposed workflow for single-cell proteomics generates results that are consistent with the known effect of Myc in changing the canonical metabolism of adult hearts. The clear separation between the cell groups in the PLS-DA analysis also suggests that the overexpression of Myc consistently induces this profile change in a majority of cardiomyocytes.

Analysis of protein changes in reference to cellular compartment increases the biological insight of single-cell proteomics

Myc promotes mitochondrial and ribosomal biogenesis, however, the dynamics of subcellular compartments and the abundance of proteins therein has not yet been studied at the single-cell level. We thus analyzed whether the single-cell proteomics data could reveal quantitative differences in subcellular compartment distributions upon Myc overexpression in a pseudobulk analysis. Therefore, we reanalyzed the data using the Systems Biology Triangle (SBT) module within iSanXoT [18–20] which analyzes changes in functional categories produced by the coordinated behaviour in abundance of the proteins within each category. We generated a protein-to-cellular compartment (q2cc) relation table by matching each identified protein with the subcellular compartment

(See figure on next page.)

Fig. 4 Pseudobulk analysis based on Zqall values, comparing control and Myc-overexpressing cardiomyocytes in the analytical dataset. **a** Violin plot representing the Zqall median values of the proteins in control and Myc-overexpressing cardiomyocytes in the analytical dataset when integrating data to all experiments in iSanXoT application. Comparison between groups was performed by Student's t-test. **b** PCA distribution of control and Myc-overexpressing cardiomyocytes based on Zqall values. PERMANOVA test results: F-value = 9.81and *p*-value = 0.001. **c** PLS-DA distribution of control and Myc-overexpressing cardiomyocytes based on Zqall values. **d** Functional protein interaction network analysis using STRING. The represented proteins are the ones extracted out of the five first components of the PLS-DA analysis with a VIP score higher or equal to 1.5. Fatty acid beta- oxidation-related proteins are highlighted in red. **e** Box plots representing the Zqall values of the fatty acid beta- oxidation-related proteins that contributes to the separation of control and Myc-overexpressing cardiomyocytes and highlighted in red in 4d. Each dot in the box plots represents the correspondent Zqall value in one cell type. A total of 81 different 10-plex TMTs were used for this analysis (analytical dataset). The final analysis includes 323 control cardiomyocytes and 324 Myc-overexpressing cardiomyocytes. Comparison between control and Myc-overexpressing groups was performed by Student's t-test. (****: $p < 0.0001$). See also Additional file 1: Fig. S5a. **f** Limma analysis is applied to evaluate the proportion of variation explained by cell size and TMT batches. As a complement of these analysis where explained variance is close to 0, *p*-value based on permutation test is computed to evaluate whether the explained proportion is significantly different from 0. In both cases, *p*-value > 0.05 (*p*-value = 1 (cell size), *p*-value = 1 (TMT))

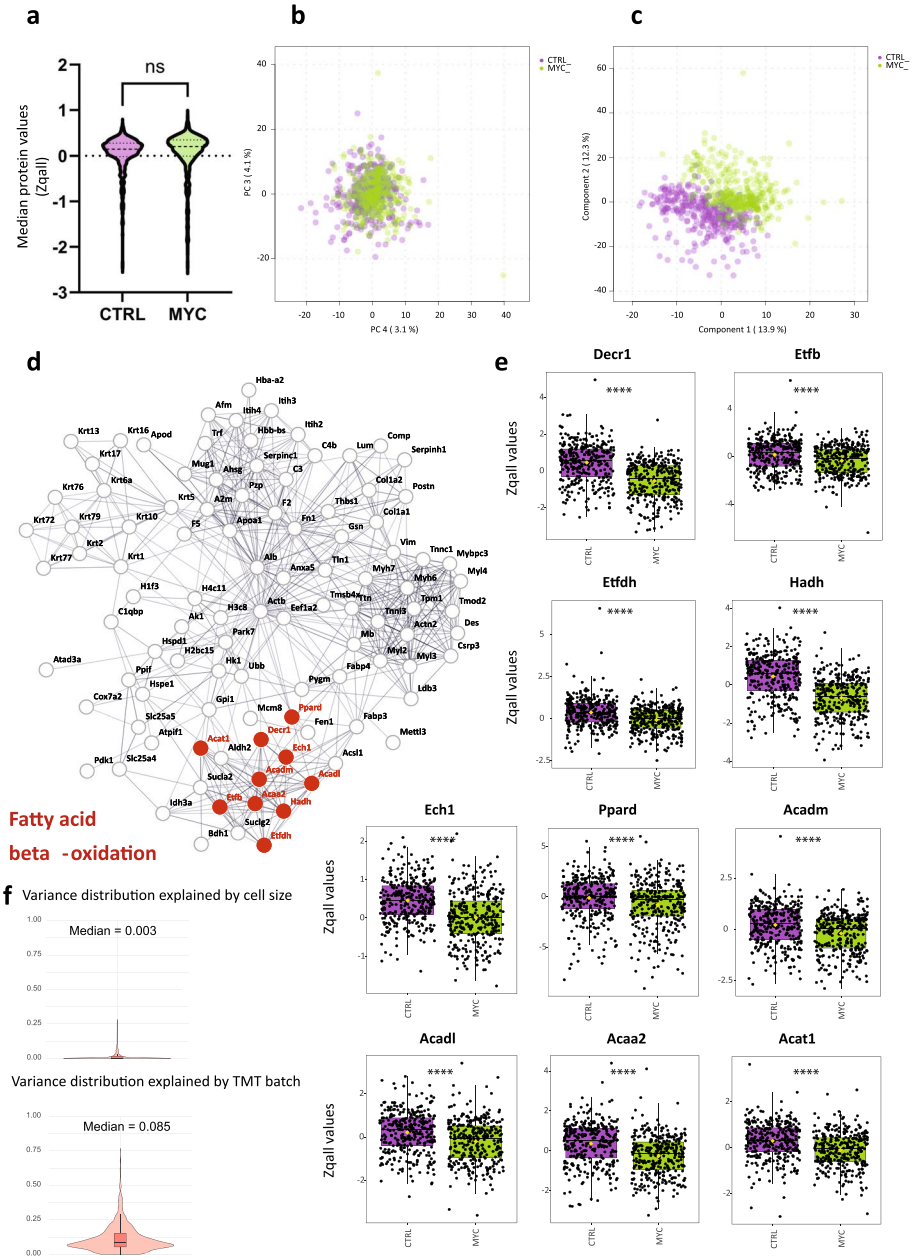

Fig. 4 (See legend on previous page.)

where it canonically localizes according to DAVID database [44, 45] (see "Methods" section), creating a total of 11 compartments, and executed the SBT module using this table (see relation table in Zenodo link [40]). This module performs, in each cell sample, GIA integrations from protein to cellular compartment and from cellular compartment to the grand mean, obtaining standardized Zqcc (e.g. quantitative protein values standardized by cellular compartment) and Zcca values (e.g. quantitative cellular compartment values standardized by cell). As expected due to standardization, no statistically significant differences were observed when comparing the median Zqcc values of control and Myc-overexpressing cardiomyocytes (Fig. 5a). However, PCA analysis using

the individual Zqcc values revealed the existence of a principal component that clearly separates control and Myc-overexpressing cardiomyocytes (Fig. 5b), being this separation considerably better than that achieved when using Zqall (e.g. quantitative protein values standarized by cell) (compare with Fig. 4b), with PERMANOVA analysis yielding improved statistics (F = 11.78, *p*-value = 0.001). These results indicate that the analysis of protein changes by cellular compartment generates a different kind of information that is more efficient in separating the two groups of cells than analyzing protein changes by cell. We then performed supervised analysis of Zqcc values with PLS-DA (Fig. 5c). In this case, again, cross-validation showed consistent R^2 and Q^2 values across components, with Q^2 exceeding 0.62 and closely tracking R^2; the Q^2 curve plateaued after two components, indicating stable predictive performance without overfitting. The classification accuracy increased smoothly from 0.785 (1 component) to 0.937 (5 components), supporting a well-generalized model (Additional file 1: Table S3). As a result of the PLS-DA analysis we extracted the proteins with highest VIP scores (123 proteins, see Zenodo link [40]). Functional analysis with these proteins also revealed that fatty acid beta-oxidation was a discriminant category, containing 13 proteins (compare Fig. 5d with Fig. 4d). As expected, Zqcc values of these proteins were decreased in a statistically significant manner in Myc-overexpressing cardiomyocytes in comparison with the controls (Fig. 5e). Cardiac muscle contraction was also highlighted as a differential category (Additional file 1: Fig. S5b). Interestingly, the analysis of Zqcc values revealed new discriminant categories that had not been detected in the Zqall analysis, like canonical glycolysis (Fig. 5d and Additional file 1: Fig. S5b). Contrary to fatty acid-beta oxidation, proteins belonging to canonical glycolysis increase significantly in Myc-overexpressing cardiomyocytes (Fig. 5f). Our data thus show that within the mitochondria there is a reduction of fatty acid beta-oxidation proteins concomitant with an increase in proteins of glycolytic metabolism in Myc-overexpressing cardiomyocytes, which correlate with the previously described metabolic changes upon Myc overexpression [23]. Overall, the results obtained by integrating protein abundance to cellular compartment indicate

(See figure on next page.)

Fig. 5 Pseudobulk analysis based on Zqcc values, comparing control and Myc-overexpressing cardiomyocytes in the analytical dataset. **a** Violin plot representing the Zqcc median values of the proteins in control and Myc-overexpressing cardiomyocytes in the analytical dataset when integrating data to the cellular compartment where they canonically belong to, using the iSanXoT application. Comparison between groups was performed by Student's t-test. **b** PCA distribution of control and Myc-overexpressing cardiomyocytes based on Zqcc values. PERMANOVA test results: F-value = 11.78 and *p*-value = 0.001. **c** PLS-DA distribution of control and Myc-overexpressing cardiomyocytes based on Zqcc values. **d** Functional protein interaction network analysis using STRING of the top proteins extracted out of the five first components of the PLS-DA analysis. The selected proteins are the ones with a VIP score higher or equal to 1.5. Fatty acid beta-oxidation-related proteins are highlighted in red. Canonical glycolysis related proteins are highlighted in blue. **e** Box plots representing the Zqcc values of the fatty acid-beta oxidation-related proteins that contributes to the separation of control and Myc-overexpressing cardiomyocytes and highlighted in red in **d**. **f** Box plots representing the Zqcc values of the canonical glycolysis-related proteins that contributes to the separation of control and Myc-overexpressing cardiomyocytes and highlighted in blue in **d**. Each dot in the box plots represents the correspondant Zqcc value in one cell type. The analysis includes 323 control cardiomyocytes and 324 Myc-overexpressing cardiomyocytes. A total of 81 different 10-plex TMTs were used for this analysis (analytical dataset). Comparison between control and Myc-overexpressing groups was performed by Student's t-test. (*: $p < 0.05$; ***: $p < 0.001$; ****: $p < 0.0001$). See also Additional file 1: Fig. S5b

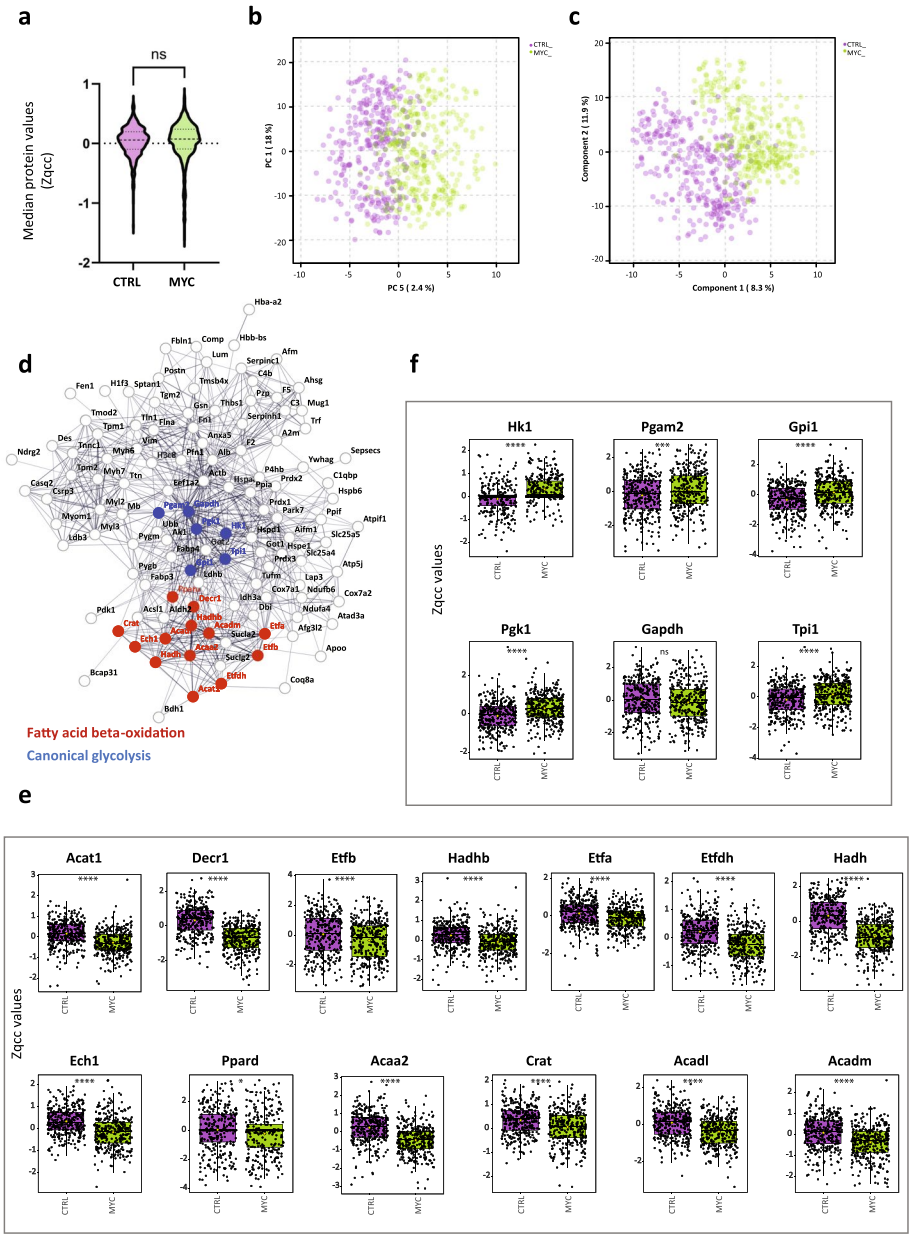

Fig. 5 (See legend on previous page.)

improved ability to discriminate changes in cellular functions, irrespectively from the global abundance of the cellular compartment.

The Zcca values obtained with the SBT algorithm collapse the information of the proteins canonically belonging to each cellular compartment into a standardized weighted average. In other words, Zcca informs about the relative abundance of each subcellular compartment as a unique entity. Among the studied compartments, endoplasmic reticulum, sarcomere, ribosome, nucleus and cytoskeleton showed a trend to increase in Myc-overexpressing with respect to control cardiomyocytes, whereas the secreted, golgi, extracellular matrix (ECM) and cytosol compartments showed the opposite trend. Mitochondria and membrane showed no

significant difference (Fig. 6). These results thus confirm that Myc-overexpressing cardiomyocytes modify the balance between different cellular compartments, with an increase in protein synthesis and maturation machinery in Myc-overexpressing

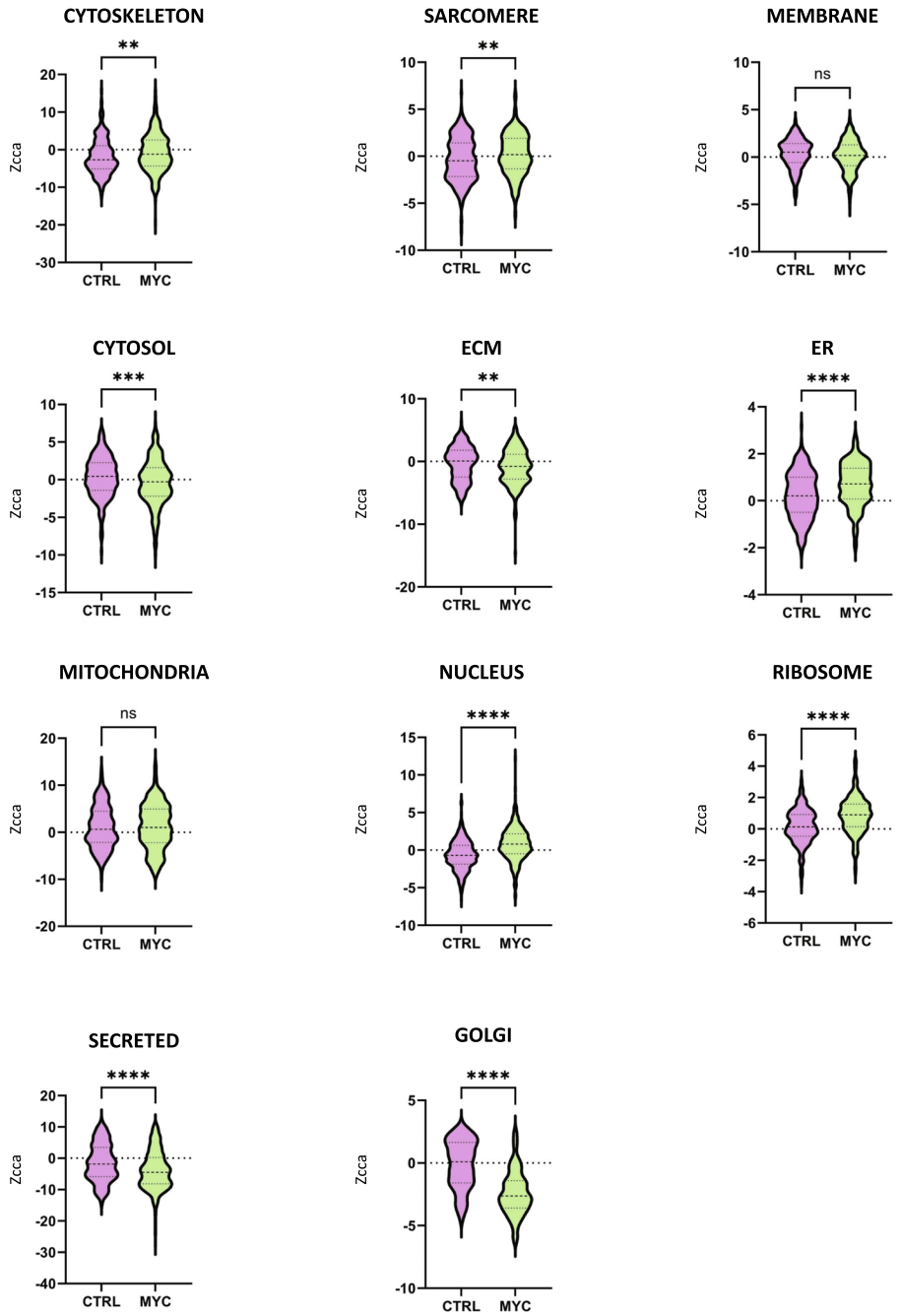

Fig. 6 Comparison of cellular compartment behaviour based on Zcca values. Violin plots representing the Zcca values per cell of each cellular compartment in control and Myc-overexpressing cardiomyocytes. Zcca value is the result of integrating cellular compartment values to the grand mean (all) using the SBT algorithm, performed with iSanXoT. The analysis includes 323 control cardiomyocytes and 324 Myc-overexpressing cardiomyocytes in each violin plot. A total of 81 different 10-plex TMTs were used for this analysis (analytical dataset). Comparison between groups was performed by Student's t-test. (ECM = Extracellular matrix; ER = Endoplasmic reticulum). (**: $p < 0.01$; ***: $p < 0.001$; ****: $p < 0.0001$). See also Additional file 1: Fig. S6

cardiomyocytes, in accordance with the known effect of Myc [46, 47] and a decrease in the secretion pathway related to golgi. Taken together, all these results show that SBT analysis offers a dual information; while Zqcc values show protein differences within each cellular compartment, Zcca offers information on the representation of the different cellular compartments within the cell. Both parameters reveal significant differences between control and Myc-overexpressing cardiomyocytes.

A cell cluster enriched in Myc-overexpressing cardiomyocytes shows markers of immaturity

Given that Zqcc and Zcca provide comprehensive and complementary information about the state of control and Myc-overexpressing cardiomyocytes, we used these two values to cluster cells following the SEURAT workflow adapted for analyzing single-cell proteomics data (see "Methods" section; and see code in Zenodo link [40]). Z values for both Zqcc and Zcca data for each cell were concatenated by row, meaning that each cell contained both protein and cellular compartment information, which was included as input for the SEURAT workflow. This joint analysis would thus help to define cell clusters based both on their cellular compartment dynamics and on the protein variations within them. Five clusters appeared in the UMAP diagram (Fig. 7a-c). Interestingly, cluster 4 was composed of 90% Myc-overexpressing cardiomyocytes, whereas the other clusters contained a balanced proportion of control and Myc-overexpressing cardiomyocytes (Fig. 7c). Cells belonging to cluster 0 exhibited higher proportion of membrane, endoplasmic reticulum, ribosome and cytoskeleton related proteins (Fig. 7e and Additional file 1: Fig. S6), which suggests elevated protein synthesis, maturation and transport. Within cellular compartments, cluster 0 shows increased electron transport chain complex I proteins or mitochondrial transporters (Lists of markers related to Fig. 7d found in Zenodo link [40] and in Additional file 1: Fig. S7), indicating a specific mitochondrial protein composition in this cluster, without significant changes in relative mitochondrial protein content.

Cluster 1 is characterized by high proportion of sarcomere and mitochondrial content (Fig. 7e and Additional file 1: Fig. S6), and by protein abundance differences within the sarcomere-cytoskeleton and the mitochondria that, together, reflect a major specialization of these subcellular components (Additional file 1: Fig. S7). Furthermore, proteins related to oxidative phosphorylation, tricarboxylic acid cycle and fatty acid beta-oxidation are high in this cluster, indicating that they contain working myocardium cardiomyocytes, with a high production of ATP based on the oxidative phosphorylation pathway. Interestingly, glycolysis and lactate biosynthetic process are increased as well, suggesting a mixed metabolic signature in parallel with the described cellular compartment alterations.

Cluster 2 is characterized by an enrichment in cell-adhesion proteins without prominent alterations in metabolic markers (Additional file 1: Fig. S7). Notably, the cells belonging to this cluster are primarily enriched in cytosolic proteins, secreted proteins and ECM, being the only cell subpopulation displaying this feature (Fig. 7e and Additional file 1: Figure S6). The coordinated change in the secretory and ECM compartments suggests that these cardiomyocytes are likely associated with the production and secretion of ECM proteins such as collagens. This evidence, together with the higher

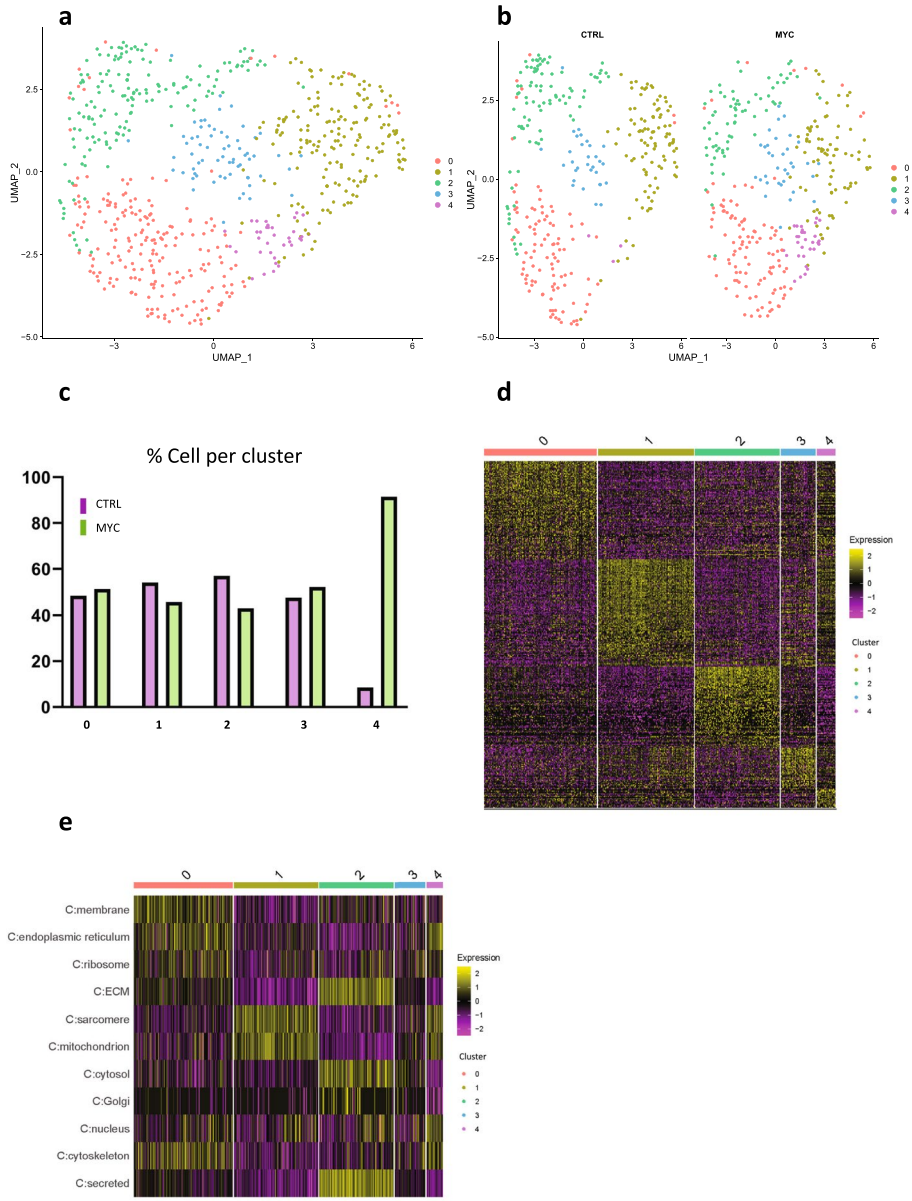

Fig. 7 Clustering analysis of control and Myc-overexpressing cardiomyocytes. **a** Results of UMAP analysis using Zqcc and Zcca values, where control and Myc-overexpressing cardiomyocytes are represented distributed in five different clusters. **b** UMAP diagrams showing control and Myc-overexpressing cardiomyocytes separately. **c** Representation of the proportion of control and Myc-overexpressing cardiomyocytes in each of the five clusters shown in Fig. 7a. (CTRL = Control cardiomyocytes; MYC = Myc-overexpressing cardiomyocytes). For this representation, the amount of each type of cardiomyocytes per cluster is normalized by the total amount of cells of each cluster. **d** Heat map showing the differential expression of the top 100 proteins contributing to the clustering of cardiomyocytes in the five clusters in UMAP shown in Fig. 7a (see also Additional file 1: Fig. S7). **e** Heat map showing the differential expression of the subcellular compartments analyzed per cluster. The analyses of all the items include 323 control cardiomyocytes and 324 Myc-overexpressing cardiomyocytes. A total of 81 different 10-plex TMTs were used for this analysis (analytical dataset). See also Additional file 1: Figs. S7, S9 and S11-S13

expression of cell-adhesion proteins suggests that these cardiomyocytes may play a structural role in the heart [48–50].

Cluster 3 displays no specific alterations in subcellular compartments, but shows the most enriched node of fatty acid beta-oxidation proteins among all the clusters

(Additional file 1: Fig. S7) indicating that it may be composed of working myocardium cardiomyocytes. The rest of metabolic markers resemble those of cluster 1, including an enrichment in glycolysis and lactate biosynthesis together with tricarboxylic acid cycle. However, no enrichment in oxidative phosphorylation proteins is observed in this cluster, suggesting a multitasking use of mitochondria (Additional file 1: Fig. S7).

Finally, cluster 4, which is predominantly composed of Myc-overexpressing cardiomyocytes, shows enrichment in proteins of glycolytic metabolism, with no altered fatty acid beta-oxidation (Additional file 1: Fig. S7). Lactate biosynthetic process, oxidative phosphorylation and tricarboxylic acid cycle appear coupled to glycolysis, suggesting that their energy production is mostly driven by the consumption of glucose. This cluster is characterized by an increase in the proportion of mitochondrial content, an aspect that has been described before in Myc-overexpressing cardiomyocytes [23, 47, 51] and that has been observed by us in P1 and adult postnatal bulk hearts (Additional file 1: Fig. S8), where it is also observed that this compartment increases in adult respect to P1 hearts. Endoplasmic reticulum, ribosomes, sarcomere-cytoskeleton or nucleus also appear as increased cellular compartments (Fig. 7e and Additional file 1: Fig. S6). These changes are consistent with an increased activity of protein synthesis, one of the Myc well-established roles [46, 47]. In contrast with the metabolic fingerprint of cluster 1 and 3, cluster 4 forms the most purely glycolytic subpopulation of cardiomyocytes given the absence of fatty acid-beta oxidation and the presence of lactate biosynthetic process together with canonical glycolysis markers among the discriminatory pathways (Additional file 1: Figs. S7 and S9).

Glycolysis and lactate biosynthesis are processes related with embryonic or early postnatal hearts [52, 53], correlating these markers with a more immature phenotype. Interestingly, Cox7a1, which is progressively activated as cardiomyocytes mature ([54] and Additional file 1: Figure S10a), represents a brake for regeneration in the zebrafish and adapts the respiratory chain to a fatty acid-based metabolism [55], is a marker for cluster 1 and not for cluster 4 (Additional file 1: Fig. S7). Similarly, Cox7a2, which is progressively repressed as cardiomyocytes mature ([54] and Additional file 1: Figure S10a) is a marker in cluster 2 (Additional file 1: Fig. S7), reinforcing the idea that this cluster does not contain fully hypertorphic working cardiomyocytes. In addition, there is a switch between Myosin 6 and 7 during postnatal heart development; while Myosin 6 is higher in adult cardiomyocytes, Myosin 7 is higher in young cardiomyocytes ([52, 53] and Additional file 1: Fig. S10b). Myosin 7 is strongly expressed specifically in cluster 4 (Additional file 1: Fig. S9d). Meanwhile, Myosin 6 is a main marker in cluster 1, suggesting again that this cluster is composed by cardiomyocytes with a mature phenotype (Additional file 1: Fig. S9d). Similarly, the relative increase in the nuclear compartment in Myc-overexpressing cardiomyocytes is consistent with previous observations in early postnatal cardiomyocytes [56] and with the increased proportion of nuclear proteins in P1 mouse hearts when compared with adult mouse hearts (Additional file 1: Fig. S10c). This is also evident when comparing control and Myc-overexpressing hearts, where the latter show a relative increase in nuclear compartment (Zcca) in both P1 and adult stages, while this value decreases with time in both heart phenotypes (Additional file 1: Fig. S8). Interestingly, it has been described that protein synthesis rates are among the lowest in mammalian adult myocardium and are downregulated compared with the

neonatal heart [57]. Our data suggests as well this fact when comparing the ribosomal levels, which are increased in Myc-overexpressing cardiomyocytes as it is seen as well in bulk analysis when comparing control and Myc-overexpressing mice hearts in P1 and adult stages. It is observed that ribosomal proteins decreases with time, however they tend to increase under Myc control (Additional file 1: Fig. S8). Overall, the subcellular and metabolic markers, together with the myosin variant profile underscore the presence of a distinct immature population promoted by the Myc transcription factor and represented in cluster 4 (Additional fie 1: Fig. S9).

To understand how Myc expression affects each of the defined clusters, we studied differential patterns at the protein and cellular compartment levels between Myc-overexpressing and control cardiomyocytes within clusters 0 to 3. Intracluster PCA and PLS-DA analysis produced a clear separation between these cell types in all clusters (Additional file 1: Fig. S11). While clusters 0–3 showed decreased markers of fatty acid beta-oxidation in Myc-overexpressing cardiomyocytes (Additional file 1: Figs. S11, S12 and S13a), clusters 1–3 showed increased markers of glycolysis (Additional file 1: Figs. S11, S12 and S13b). Moreover, nuclear and ribosomal compartments are increased in Myc when compared to control cardiomyocytes in most clusters (Additional file 1: Fig. S13d). Data also show that Myosin 7 tends to increase in the four clusters in Myc-overexpressing cardiomyocytes with respect to control ones, whereas Myosin 6 only increases mildly in clusters 0 and 3 (Additional file 1: Fig. S13c). These results suggest that Myc promotes an immature signature in all cardiomyocyte subtypes but with different efficacies. They also indicate that Myc expands a minoritary subtype of cardiomyocytes represented by cluster 4, or is able to reprogram cardiomyocytes from other clusters into the cluster 4 phenotype.

Analysis of the total cellular protein content suggested that cells in clusters 0 and 4 have smaller size (Additional file 1: Figs. S14c and S15a, b). The smaller size is more prominent in Myc-overexpressing cardiomyocytes as previously described (Additional file 1: Fig. S16a-c) and see references [24] and [25]. Recent studies have shown a negative correlation between ploidy and the regenerative capacity of the heart [58–60]. It has been described that cellular protein content scales sublinearly with ploidy [61] and histone amount correlates with ploidy [62]. To compare cellular with histone content we generated a chromatin category composed by the histones identified in the analysis (Additional file 1: Table S4) and used iSanXoT to perform the protein to category integration. We found that the protein content of the chromatin category had the same behaviour across the clusters than that of the whole cell defined by the total cellular protein content (Additional file 1: Fig. S14c,d). Indeed the total cellular protein content correlated very well with the chromatin content (Additional file 1: Fig. S17a), in agreement with the known correlation between ploidy and cell size. We also found that the average chromatin protein content of clusters 0 and 4 was approximately two-fold lower than the rest of clusters, suggesting that clusters 0 and 4 had a lower ploidy (Additional file 1: Fig. S15d). The same decrease was observed when comparing the average total cell protein content (Additional file 1: Fig. S15b). Concerning the rest of cellular compartments, as a general trend, the distribution among clusters of their protein content followed the same trend than the total cellular content (Additional file 1: Fig. S14e-o) and showed a very clear correlation with protein content of chromatin (Additional file 1: Fig. S17b-l),

corroborating that ploidy controls the scaling of cellular size and organelles [61, 63–65]. Our data agree with previous reports describing that Myc-overexpression promotes lower ploidy in cardiomyocytes [25], which correlates with an improved recovery after infarction. These results further support the specific effects of Myc in expanding cluster 4 cardiomyocytes and highlight their potential pro-regenerative fingerprint.

Our study underscores the relevance of studying non-normalized protein contents of total-cell and subcellular-compartments, the analysis of the relative abundance of each subcellular compartment within the cell and the relative abundance of proteins within their specific cellular compartment. Each of these levels provides different and complementary information that allows to better understand cellular dynamics. While global changes may mask subtle changes at the subcellular level, one should not disregard all levels of information for better profiling a cell signature. Overall, the integration of information from different layers of proteomics data provided by the iSanXoT framework allowed to characterize single-cardiomyocytes with an unprecedented subcellular detail. Using this approach, our study identified a minoritary cardiomyocyte population with pro-regenerative characteristics and showed its expansion upon Myc-overexpression.

Discussion

In this work we show that the analysis of single-cell proteomics data may be improved using the integrative statistical framework provided by iSanXoT. This was done using a cell population that contains control and Myc-overexpressing mouse cardiomyocytes. While it is known that Myc overexpression promotes a more immature metabolism in adult cardiomyocytes [23], this metabolic fingerprint has never been studied before at the single-cardiomyocyte level. Working with single-adult cardiomyocytes is a challenging task due to the difficulty in handling these cells and also in regard to the high dynamic range exhibited by this cell type at the proteome level. Technically, our work first highlights the importance of using an appropriate method to isolate individual cardiomyocytes for conducting single-cell proteomics. Although current publications in the field make use of advanced automatized cell isolation strategies for analyzing cardiomyocytes, some of them compatible with sample preparation for MS analyses [16, 26], here we intend to highlight the need of a careful control of cell integrity when isolating single-adult cardiomyocytes. These cells require a step-by-step controlled procedure that preserves the fitness of the isolated cells to obtain robust quantitative data.

Our results also show the ability of iSanXoT to remove batch- and cell size- effects in single-cell proteomics data, avoiding the missinterpretation of information biased by cell size. Furthermore, cellular compartment-based normalization strategies have not been explored in the field so far. The flexibility offered by iSanXoT allows the use of the SBT algorithm for straightforward analysis of single-cell alterations based on protein abundance within cellular compartments, and to study global changes of the compartments at the single-cell level. Of note, using a statistically-validated algorithm to integrate the quantitative protein information at the subcellular compartment level is a new concept in single-cell proteomics. We should note here that the canonical localization of proteins in the correspondant subcellular compartment is limited by the database knowledge and therefore may carry possible miss-assignements of some proteins in a specific context, for instance of proteins that traslocate to different cellular compartments.

Our data demonstrate that integration of protein information to the cellular compartment using the SBT algorithm in pseudobulk analysis detects the known effect of Myc overexpression in increasing ribosomal protein [46, 47], but also the down-regulation of fatty acid beta-oxidation and the up-regulation of glycolysis [23], which were otherwise not detected. This metabolic fingerprint can be correlated with a pro-regenerative capacity in Myc-overexpressing cells, since it has been described that the inhibition of fatty acid beta-oxidation reactivates the regenerative capacity of the heart [66, 67] and at the same time the increase in glycolysis has been correlated with a higher regenerative capacity in Zebrafish heart [68].

Interestingly, our results show that using information about subcellular compartment protein abundance provides improved cell classification and a highly valuable insight into the biology of clustered cell subgroups. In our case, a cluster enriched in Myc-overexpressing cardiomyocytes is shown to contain cells with increased proportions of mitochondrial and ribosomal content, as described for cells overexpressing Myc [23, 46, 47, 51]. Nuclear compartment is linked as well to this cluster of Myc-overexpressing cardiomyocytes. Interestingly, cluster 2 shows larger proportions of ECM, golgi and secretory pathway compartments, linking this group of cells with a role in ECM secretion, which has been connected to the golgi activity [69].

The clustering analysis indicated that all except one cluster show a similar proportion of control and Myc-overexpressing cardiomyocytes, however, intracluster analyses reveal that Myc overexpression promotes a metabolic switch to immaturity. This fact may underlie the contradicting result that for several clusters markers of fatty acid oxidation coexist with markers of glycolysis. Nonetheless, not all clusters are equally sensitive to Myc overexpression, with cluster 0 apparently being more resistant to it, even though Myc levels in the studied model have been characterized to be within physiological levels, corresponding to one extra allele of Myc homogeneously expressed in the nucleus along all the Myc-overexpressing cardiomyocytes [24, 25].

Moreover, the clustering analysis showed that the cell cluster highly enriched in Myc-overexpressing cardiomyocytes is characterized by canonical glycolysis as one of the most represented pathways, together with lactate biosynthesis and the absence of markers of fatty acid beta-oxidation. The prevalence of fetal Myosin 7 over the adult Myosin 6 in this cluster points as well to the immaturity of this group of cells. Moreover, the lower DNA content shown by this group of cells correlates as well them with a pro-regenerative phenotype [25, 58–60]. Two possible mechanisms can be envisioned to explain the expansion of cluster 4 by Myc; expansion of the naturally scarce population of cluster 4 cardiomyocytes, or conversion of cardiomyocytes originally belonging to other clusters. In the case of conversion, given that ploidy reduction is unlikely to occur in cardiomyocytes [70], the cluster 0 cardiomyocytes would be the most likely source, provided their similar ploidy to cardiomyocytes in cluster 4. The scarcity of this type of cardiomyocytes in wild type hearts may underlie the poor regenerative ability of the adult mammalian heart and our results suggest the ability of Myc to specifically expand this rare population. Our study thus suggests different sensitivity of cardiomyocyte populations to Myc overexpression and specific expansion of a rare, potentially pro-regenerative cardiomyocyte population.

Overall, the methods proposed in this work allow profiling specific cardiomyocyte phenotypes within the heart, providing different layers of information that can help to better understand the nature of these cardiomyocyte subgroups and their response to regenerative therapies. The analytical details and precision offered by this procedure can be of substantial interest for defining molecular phenotypes difficult to be detected by current methodologies and may therefore be of general utility for single-cell proteomics analyses.

Conclusions

In this work we show that using optimized protocols it is possible to study cardiomyocyte subpopulations by single-cell proteomics. We also show that by using integrative quantitative workflows to process single-cell data it is possible to add additional layers of information to study protein alterations within cellular compartments or the behaviour of the compartments themselves; these are unconventional pieces of information that may help to interpret these complex datasets.

Heart regeneration is a hot topic in cardiology and it is known that overexpression of transcription factor Myc enhances cardiomyocyte proliferative capacity. However, it was unknown whether all cardiomyocytes had the same response to Myc overexpression since it had not been studied yet at the single-cell level. Here we show that the individual cardiomyocyte response to Myc is heterogeneous and promotes different degrees of immaturity expanding a subpopulation of purely immature pro-regenerative cardiomyocytes, highlighting the relevance of conducting studies at the single-cell level when evaluating regenerative therapies. Beyond basic research, this new approach will allow the deep phenotypation of cardiomyopathies in humans.

Methods

Animal models

Myc-overexpressing mice lines have been previously described and characterized, where Myc transcription factor was knocked-in in the locus of Rosa26 allowing an homogeneous expression of this transcription factor in the nucleus of cardiomyocytes [24, 32, 71, 72]. Mice were genotyped by PCR. All experiments were conducted using male wild-type mice, aged over 14 weeks for adults and 1 day (P1) for neonatal mice. Cryo-sections of control and Myc-overexpressing hearts were evaluated by optical microscopy (Leica DM6B Upright Navigator).

Single-cardiomyocyte isolation

For isolating single-cardiomyocytes, two methods were followed: a manual isolation genuinely designed in our lab for isolating single- cardiomyocytes, and a regular fluorescence-activated cell sorting (FACS) isolation. For both methods, mice hearts ($n=2$ of each mice type) were dissected and conveniently prepared for Langendorff perfusion after excluding atria [73]. In the manual isolation [33, 36], cardiomyocytes obtained after Langendorff perfusion were plated in a 9 cm dish containing 1 × PBS. Subsequently, cardiomyocytes were handled with a steripette and individually distributed in 1 × PBS droplets under the scope. Cell viability was checked in each droplet under the optical microscope and individual cells were picked up with a regular pippete in 5 µl 1 × PBS and disposed in a 96 well-plate

conveniently refrigerated at 4ºC. This procedure ensured single-cells were collected and no doublets were disposed into each well. Each well contained a volume of 15 µl HPLC-grade water (Sigma) for initiating the cell lysis after cell disposal. In the FACS-sorting assisted isolation method [39], cardiomyocytes resultant from the Langendorff procedure were disposed in an Eppendorf tube containing 1 × PBS. BD FACSAria Fusion cell sorter was used for isolating single-cardiomyocytes using the maximum 100 µm nozzle for sheath pressure. Convenient gating was designed for excluding doublets. Single- cardiomyocytes were disposed in a volume of 5 µl 1 × PBS into each 96-well plate conveniently refrigerated at 4ºC and containing 15 µl HPLC-grade water (Sigma) for initiating the cell lysis. Samples were frozen at −80ºC after manual and FACS sorting collection. 12 different 10-plex TMTs were used in the pilot analysis [33]; in the comparison of FACS (dataset 2) and manual isolation (dataset3) analyses, 14 different 10-plex TMTs were used [36, 39]. The 14 different 10-plex TMTs analyzed in the manual isolation section (Fig. 3) were included in the analytical dataset where a total of 81 TMTs were included.

Preparation of single-cell samples for mass spectrometry analysis

Cells lysates were thawed and prepared for mass spectrometry analysis equally in manually- and FACS-sorted collected cells as described in [74]. Scope-MS protocol was followed for sample preparation [4]. Specifically, samples were sonicated for 3 min in a water bath sonicator (Branson 3800 ultrasonic bath) and subsequently denatured under 95ºC for 15 min. Proteins were digested overnight with trypsin at 37ºC. Per cell, 75 ng of trypsin were added. After digestion, samples were labelled for 1 h with 1 ul of 85 mM 10 plex-Tandem Mass Tag (TMT) isobaric label reagents for multiplexing [4, 5, 10]. Tag 130 N was left empty and tag 131 was used for booster cells containing a lysate correspondent to 200 cardiomyocytes composed by 100 control cardiomyocytes and 100 Myc-overexpressing cardiomyocytes [74]. Cells for booster channel were prepared following the same procedure as for single cells, but adjusting trypsin values to 150 ng for digestion. Final concentration of 1.2% of Tri-fluoroacetic acid (TFA) was used for quenching the labelling reaction. Each TMT-single-cell-labelled batch was composed by four control cardiomyocytes and four Myc-overexpressing cardiomyocytes. For labelling, two designs were followed so that each type of cardiomyocyte (control or Myc-overexpressing cardiomyocytes) was differentially labelled with alternative tags to avoid a biased influenced by tag type (Fig. 1c). After labelling, samples were pooled as recommended by manufacturer and desalted on C18 Omix cartridges (Agilent Technologies) [74] (Fig. 1b). Eluted peptides were concentrated till dryness and resuspended in 0.1% formic acid in 2% acetonitrile (ACN) for mass spectrometry (MS) analysis.

Preparation of bulk samples for mass spectrometry analysis

Mice hearts were dissected and disrupted with Fastprep-24 ™ 5G (MP Biomedicals) in a lysis buffer containing 2% SDS [75, 76]. Samples were digested with trypsin (Promega) overnight in Pall filters (Pall Corporation) following the FASP protocol [77] and using a 1:30 ratio of trypsin (1 ug trypsin per 30 ug of protein) after reducing and alkylating with dithiothreitol (50 mM) (DTT, Sigma) and iodoacetamide (20 mM) (IAA, Sigma). The resulting peptides were desalted on C18 Oasis cartridges (Waters Corporation) and vacuum dried. The peptides were then subjected to isobaric TMT labelling following

manufacturer's instructions. An internal control was constructed by pooling all the samples and using it as an internal reference. After labelling, samples were desalted on C18 Oasis cartridges. Eluted and desalted peptides were concentrated till dryness. Samples were re-constituted in 0.1% formic acid in 2% ACN for individual MS analysis. In the case of the bulk analysis where we compared control and Myc-overexpressing hearts, eluted peptides were submitted to a high pH fractionation procedure following manufacturer´s instructions (High pH Reversed phase-peptide fractionation kit-Pierce). Eluted peptides from each fraction were concentrated till dryness and re-constituted in 0.1% formic acid in 2% ACN for individual MS analysis.

LC–MS data collection of single-cell samples

Peptides were loaded onto a 2 cm C18 trap column (ThermoFisher Scientific) connected in-line to a 50 cm C18 reversed-phase analytical column (ThermoFisher Scientific) with 100% buffer A (0.1% formic acid in water). Vanquish UPLC system (dataset 1, [33]) (ThermoFisher Scientific) or Ultimate 3000 nano UPLC system (Analytical dataset [36] and datasets 2 [39], for FACS isolation and 3 [36], for manual isolation) (Thermo Fisher Scientific) were used coupled via a nanoelectrospray ion source (ThermoFisher Scientific) to an Orbitrap Eclipse tribrid mass spectrometer equipped with a FAIMS Pro device (ThermoFisher Scientific). Peptides were eluted over a 100 min linear gradient of buffer B (0.1% formic acid in 80% ACN) at 200 nL/min. FAIMS was configured to alternate between compensation voltages of −50 V and −70 V across two consecutive cycles. Data were acquired following the real-time search-assisted acquisition RETICLE method [78]. In this method, MS1 spectra were acquired at 120,000 resolution with a scan range from 375 to 1500 m/*z*, normalized AGC target of 300% and maximum injection time of 50 ms. Precursors were filtered with a dynamic exclusion of 120 s and were isolated for MS2 analysis in the quadrupole with a 0.7 Th window. Fragmentation was performed at linear ion trap (LIT) with 30 normalized collision energy by collision induced dissociation (CID) and MS2 spectra were acquired in the LIT MS2 at scan rate turbo. LIT MS2 spectra were subjected to real time search (RTS) using the mouse database obtained from Swiss-Prot with isoforms (downloaded on 05/2021) and trypsin set as enzyme. TMT 10-plex (+229.1629 Da) on lysine (K) and N-termini was set as static modification, and oxidation of methionine (M) as variable modification. Maximum missed cleavages were set to 1 and maximum variable modifications was set to 1. FDR filtering was not enabled, maximum search time was set to 40 ms and the scoring threshold was set to 1.4 Xcorr, 0.1 dCN, and 10 ppm precursor tolerance. *Use as Trigger Only* was checked and *Close-Out* was enabled with maximum number of peptides per protein set to 4. Subsequently, precursors identified via RTS were isolated in the quadrupole with a 0.7 Th window, ions were collected for a maximum injection time of 500 ms and normalized AGC target of 500%, fragmented with 38 normalized collision energy by higher energy collision dissociation (HCD) and MS2 spectra were acquired in the Orbitrap at 120,000 resolution with first mass set to 110 m/*z* [78]. MS instrument stability was checked by monitoring the mass accuracy of the lock mass 445.12 every second to ensure homogenous performance of the system for all TMT batches. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via

the PRIDE [79, 80] partner repository with the dataset identifiers listed along the text and in the Availability of Data and Materials section.

LC–MS data collection of bulk samples

Bulk samples were acquired using an Ultimate 3000 nano UPLC system (Thermo Fisher Scientific) coupled via a nanoelectrospray ion source (Thermo Fisher Scientific) to a Q Exactive HF mass spectrometer (Thermo Fisher Scientific). C18-based reversed-phase separation was used with a 2-cm trap column connected in-line to a 50-cm analytical column (Thermo Fisher Scientific). Peptides were loaded in 100% buffer A (0.1% formic acid in water) and eluted with a 300 min linear gradient of buffer B (0.1% formic acid in 80% ACN) at 200 nL/min. Mass spectra were acquired in a data-dependent manner, with an automatic switch between MS and MS/MS using a top 15 method. MS spectra were acquired in the Orbitrap analyzer with a mass range of 390–1700 m/z and 60,000 resolution, using an AGC of 3e6 and a maximum injection time of 50 ms. HCD fragmentation was performed at 30 normalized HCD and MS/MS spectra were analyzed at 30,000 resolution in the Orbitrap, using an AGC of 2e5 and a maximum injection time of 120 ms. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [79, 80] partner repository with the dataset identifiers listed in the Availability of Data and Materials Section [75, 76].

Data analysis of single-cell and bulk samples

Single-cell and bulk raw files were submitted to Proteome Discoverer 2.5 (Thermo Fisher Scientific) with the built-in TMT Reporter ion quantification workflow for obtaining reporter ion quantification values. Default settings were applied, with trypsin as enzyme specificity. Spectra were matched against the 10090 *Mus musculus* database obtained from Uniprot (Swiss-Prot with isoforms for single-cell data and Swiss-Prot and TrEMBL with isoforms for bulk analysis (downloaded on 05/2021)). Oxidation (M) and acetyl on protein N-termini were set as dynamic modifications. TMT-10 plex (+229.163 Da) as tag on both peptide N-termini and Lys (K) residues was set as static modification. Spectra were searched using the Sequest HT search engine with 800 ppm of precursor mass tolerance, followed by post-processing where the false discovery rate (FDR) was set to 1% [81], with an additional filter for precursor mass tolerance of 15 ppm [82]. For the reporter ion quantification, normalization mode and scaling mode were set to None and average reporter s/n threshold was set to 0. Isotopic error correction was applied including the correspondent batch correction matrix. PSM txt files output from Proteome Discoverer were analyzed in iSanXoT [18] for quantitative and statistical analysis. Quantitative values (referred to as X values according to iSanXoT nomenclature) were expressed as the 2-base logarithm of the ratio of the reporter intensity to that of an internal standard composed by a pool of all the analyzed samples, in the case of bulk data analysis, or to the average of the reporter intensities of all individual cells within each TMT, in the case of single-cell data analysis. In the timeline bulk analysis we quantified a total of 3400 proteins and 1724 out of them were identified with 2 or more peptides. In the bulk analysis where we compared control and Myc-overexpressing hearts we quantified a total of 5194 proteins and 3225 of them were quantified with 2 or more peptides.

Quantitative analysis of single-cell data

iSanXoT calibrates the experimental variance of the input data at the scan level [18, 35], and the GIA algorithm within iSanXoT automatically models quantification errors and expresses quantitative values in terms of standardized log2-ratios (Z values) [20]. Briefly, iSanXoT performs several steps of iterative modelling. The first one, called variance calibration, fits the variance distribution of the quantitative data to reporter intensities. Once variance is calibrated, iSanXoT computes, in each cell, weighted averages to *integrate* quantitative values at the reporter (or scan) level to obtain peptide values. Similarly, peptide values are integrated to obtain protein values, and protein values are integrated to a grand mean per each cell. Standardized Z values are computed at each of the levels (scan, peptide and protein) so that the transformed data are expected to follow a standard distribution. Thus, in the case of proteins, quantitative protein values are standardized by cell content (grand mean). In each integration step the variance distribution is iteratively modeled by fitting the Z values to the expected N(0,1) sigmoidal curve, obtaining a residual variance specific of each integration. iSanXoT automatically lists residual variances and generates plots to check the accuracy of the iterative modelling at each step and for each cell. Representative examples of calibration and integration plots and the distribution of residual variances are shown in Additional file 1: Fig. S4.

To generate the protein to cellular compartment relation table, a reference list of proteins extracted out of the timelapse bulk analysis containing all the identified proteins in single-cell data has been used as a reference list. In this list, the proteins were matched to the canonical celular compartment where they are localized according to DAVID database [44, 45]. Later, a manual curation of gene ontology terms, using information from Uniprot database and in some specific cases from the literature, allowed us to compile the information into 11 compartments. The table with the proteins and their putative compartment assignations is included in Zenodo link [40].

The SBT module of iSanXoT [18, 20] was used to integrate quantitative protein values to the cellular compartment where each protein was assigned, obtaining the protein value data standardized by cellular compartment (Zqcc). Subsequently, cellular compartment values were integrated to the grand mean, obtaining cellular compartment values standardized by cell (Zcca). For the analysis of cellular compartment and total cell protein abundances, Xa and Xcc values obtained from iSanXoT were used, respectively, where Xcc is the weighted average of the quantitative protein values within each compartment and Xa is the weighted average of all the cell compartments (Xcc) values. Note that these X values are the ones obtained before standardization and hence are not normalized. Data were imputed using a constant value [83]. Interaction network and enrichment analysis were performed using STRING [84]. For comparing trends in the analytical pseudobulk strategy, median values of proteins present in at least 30% of cells were considered. Linear multivariate analyses such as Principal Component Analysis (PCA) and Partial Least Square Discriminant Analysis (PLS-DA) were conducted with Metaboanalyst (vs. 6.0) [85, 86] while non-linear multivariate approaches, such as Uniform Manifold Approximation and Projection (UMAP) is performed within SEURAT (vs. 4.3.0) workflow [87]. In Metaboanalyst, we applied filters based on standard deviation and on mean quantitative values; since the Z data was standardized, no normalization was required. Top ranked proteins with VIP score higher than 1.5 in the PLS-DA

analysis were selected for functional analysis among the first five components. Permutational Multivariate Analysis of Variance (PERMANOVA), a non-parametric multivariate analysis of variance based on permutation tests was applied for assessing group separation in the PCA analyses following an R code [88, 89]. As a result F-value, a measurement of the between-group variance and *p*-value, a measurement of the significance of the between-groups separation is provided. The contribution of experimental covariates (TMT set and cell size) to the observed variance was performed by using the Limma package in R. For each feature we estimated the coefficient and residual variance across samples, and computed the proportion of variance attributable to each covariate. Proportions were then summarized across all features to obtain an overall distribution of variance explained across samples [90–92]. To evaluate the contribution of the covariates to the total variance in the proteomic dataset a linear model fitted per protein was used. To determine whether this proportion was significantly different from zero, a permutation test was performed, where the values of the covariate were randomly permuted across samples for 1000 times and the proportion of variance explained was recalculated for each permutation. The *p*-value was estimated as the fraction of permutations that produced a proportion of variance greater or equal to the observed value. PLS-DA models were evaluated using crossvalidation analysis in Metaboanalyst (vs 6.0) to assess model performance together with its prediction capacity [93, 94]. For finding markers in the clustering analysis, ROC analysis was used and the positive elements with area under the curve (AUC) value higher than 0.6 were selected. In SEURAT, the data generated by iSanXoT were included in the scale step of SEURAT, where each included feature is scaled along the total number of cells. Both Zqcc and Zcca values are concatenated by row in the same input file for SEURAT. Chi-square test was used to evaluate the dependency of the clusterization in UMAP with TMT batch. The used codes are available in the provided zenodo links [40, 95, 96].

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-026-04110-1.

Additional file 1: Fig. S1. PCA of single-cell proteomics data normalized by different criteria. Fig. S2. PCA of non-normalized single-cell proteomics data. Fig. S3. Normalizing strategies for analyzing mass spectrometry-based single-cell proteomics data. Fig. S4. Performance of iSanXoT for modelling single-cell proteomics data. Fig. S5. Functional analysis of proteins discriminating between control and Myc-overexpressing cardiomyocytes. Fig. S6. Variation of Zcca values across clusters. Fig. S7. Interactome network of cluster markers defined by UMAP analysis. Fig. S8. Evaluation of cellular compartments in P1 and adult stages of Myc-overexpressing hearts respect to control hearts. Fig. S9. Relative abundances of selected protein markers related to immaturity found in the clustering analyses Fig. S10. Abundance levels of Myosins, Cox7a and nuclear proteins in bulk mice heart analyses. Fig. S11. Multivariate analysis showing the intracluster separation of control and Myc-overexpressing cardiomyocytes. Fig. S12. Proteins discriminating control and Myc-overexpressing cardiomyocytes within each cluster based on the PLS-DA analysis shown in Fig. S11. Fig. S13. Dot plot representing the differential abundance of protein and subcellular compartments within each cluster in the subpopulation of Myc-overexpressing cardiomyocytes respect to control ones. Fig. S14. Distribution of total cell and of cellular compartment protein abundances across clusters. Fig. S15. Detailed representation of cell size and chromatin variation along clusters. Fig. S16. Distribution of total cell protein abundances distinguishing control and Myc-overexpressing cardiomyocytes. Fig. S17. Correlation between total cell or cellular compartment protein abundances vs chromatin protein abundances. Table S1. QC metrics comparing FACS and MANUAL isolation methods for single-cardiomyocytes. Table S2. PLS-DA cross validation details for Fig. 4c. Table S3. PLS-DA cross validation details for Fig. 5c. Table S4. Proteins included in the chromatin category. Table S5. Chi-squared analysis of the TMT batch effect in clusterization of data.

Acknowledgments
We acknowledge Eneritz Agirre (Karolinska Institutet, Sweden) and the Bioinformatic Unit (CNIC, Spain) for the assistance in the bioinformatic analyses. We also acknowledge the Animal Facility, Flow Cytometry Facility and Microscopy Facility (CNIC, Spain). We acknowledge the Leducq Fondation funded Consortium "Redox Regulation of Cardiomyocyte Renewal" TNE-17CVD04, for scientific feedback and funding. We acknowledge the technical assistance of ThermoFisher Scientific scientists in Bremen.

Peer review information
Fengchao Yu and Tim Sands were the primary editors of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team. The peer-review history is available in the online version of this article.

Authors' contributions
M.T., J.V., C.M.V., C.V.dC. and E.C. designed the experiments. C.V.dC., R.S., S.M.S. and C.M.V. isolated the cardiomyocytes, contributed with optical microscopy and the correspondant data analysis. C.M.V. carried out experiments and conducted MS acquisition and analyses. J.V., M.T. and C.M.V. wrote the manuscript. J.V., J.M.R., C.M.V., M.T. and E.C. designed and implemented the computational workflow. J.V. and J.M.R. implemented the iSanXot workflow. C.T. assisted with computational supervision. A.V., R.Z. and E.C. assisted with the MS analysis, helped develop the instrument methods, and contributed to the overall functioning of the method. All authors proofread and contributed to the manuscript.

Funding
This work was supported by the Leducq Fondation funded Consortium "Redox Regulation of Cardiomyocyte Renewal" TNE-17CVD04, by the competitive grants from the Spanish Ministry of Science, Innovation and Universities (PGC2018-097019-B-I00, PID2021-122348NB-I00, PLEC2022-009235 and PLEC2022-009298), PID2024-155650NB-I00 financed by MICIU/AEI /10.13039/501100011033 and by FEDER, UE. It has also been funded by the Instituto de Salud Carlos III (Fondo de Investigación Sanitaria grant PRB3 (PT17/0019/0003- ISCIII-SGEFI / ERDF, ProteoRed), Comunidad de Madrid (IMMUNO-VAR, P2022/BMD-7333), "la Caixa" Banking Foundation (project codes HR17-00247 and HR22-00253), European Commission H2020 Program grant SC1-BHC-07-2019. Ref. 874764 "REANIMA", Comunidad de Madrid grant P2022/BMD-7245 and Funded by the European Union (ERC AdG REACTIVA Ref. 101142005). The CNIC is supported by the Instituto de Salud Carlos III (ISCIII), the Ministerio de Ciencia, Innovación y Universidades (MICIU) and the Pro CNIC Foundation, and is a Severo Ochoa Center of Excellence (grant CEX2020-001041-S funded by MICIU/AEI/10.13039/501100011033).

Data availability
The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [79, 80] partner repository with the dataset identifier PXD064499 and https://doi.org/10.6019/PXD064499 for the dataset 1 (pilot dataset) [33], PXD064501 and https://doi.org/10.6019/PXD064501 for the FACS dataset (dataset2) [39] and PXD064518 and https://doi.org/10.6019/PXD064518 for the Analytical dataset [36]. Notice that TMT samples 5–9 and 12–20 within the Analytical Dataset correspond to the dataset 3 (manual isolation). PXD064994 and https://doi.org/10.6019/PXD064994 (timeline analysis) [75] and PXD073496 and https://doi.org/10.6019/PXD073496 (Myc-overexpressing mouse heart analysis) [76]. The manuscript includes information as complementary data that has been referenced along the manuscript and can be found in the zenodo link https://doi.org/10.5281/zenodo.18379944 [40]. Extra Zenodo links are included with information related to each of the figures: https://doi.org/10.5281/zenodo.19484833: Figs. 1, 2, 3, S1, S2, S3 [95]. https://doi.org/10.5281/zenodo.20273639: Figs. 4, 5, 6, 7 [96]. https://doi.org/10.5281/zenodo.18375680: Figures S4, S5, S6, S7, S8, S9, S10 [97]. https://doi.org/10.5281/zenodo.18375720: Figures S11, S12, S13, S14, S15, S16, S17 [98]. To support the described Single-Cell Proteomics workflow, we have provided an iSanXoT workflow template that generates the required integrations. This template includes the task tables necessary for each involved iSanXoT module. The iSanXoT workflow template, along with the required input files for executing this Single-Cell Proteomics workflow can be found in Zenodo link [99]. For detailed instructions on importing the workflow template, refer to the iSanXoT documentation at https://cnic-proteomics.github.io/iSanXoT/ [100].

Declarations

Ethics approval and consent to participate
All mice experiments were maintained and handled according to the recommendations of the CNIC Ethics Committee, the Spanish laws and the EU Directive 2010/63/EU for the use of animals in research. We have complied with all relevant ethical regulations for animal use. The experimental protocols involving animals were approved by the CNIC and Universidad Autónoma de Madrid Committees for "Ética y Bienestar Animal" and the area of "Protección Animal" of the Comunidad de Madrid with reference PROEX 220/15.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 29 August 2025 Accepted: 11 May 2026
Published online: 22 May 2026

References

1. Aldridge S, Teichmann SA. Single cell transcriptomics comes of age. Nat Commun. 2020;11(1):4307.
2. Gulati GS, D'Silva JP, Liu Y, Wang L, Newman AM. Profiling cell identity and tissue architecture with single-cell and spatial transcriptomics. Nat Rev Mol Cell Biol. 2024. https://doi.org/10.1038/s41580-024-00768-2.
3. Stegle O, Teichmann SA, Marioni JC. Computational and analytical challenges in single-cell transcriptomics. Nat Rev Genet. 2015;16(3):133–45.
4. Budnik B, Levy E, Harmange G, Slavov N. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. Genome Biol. 2018;19(1):161.
5. Petelski AA, Emmott E, Leduc A, Huffman RG, Specht H, Perlman DH, et al. Multiplexed single-cell proteomics using SCoPE2. Nat Protoc. 2021;16(12):5398–425.
6. Schoof EM, Furtwängler B, Üresin N, Rapin N, Savickas S, Gentil C, et al. Quantitative single-cell proteomics as a tool to characterize cellular hierarchies. Nat Commun. 2021;12(1):3341.
7. Petrosius V, Aragon-Fernandez P, Üresin N, Kovacs G, Phlairaharn T, Furtwängler B, et al. Exploration of cell state heterogeneity using single-cell proteomics through sensitivity-tailored data-independent acquisition. Nat Commun. 2023;14(1):5910.
8. Petrosius V, Aragon-Fernandez P, Arrey TN, Woessman J, Üresin N, de Boer B, et al. Quantitative label-free single-cell proteomics on the Orbitrap Astral MS. Mol Cell Proteomics. 2025. https://doi.org/10.1016/j.mcpro.2025.100982.
9. Johnston SM, Webber KGI, Xie X, Truong T, Nydegger A, Lin H-JL, et al. Rapid, One-Step Sample Processing for Label-Free Single-Cell Proteomics. J Am Soc Mass Spectrom. 2023;34(8):1701–7.
10 Specht H, Emmott E, Petelski AA, Huffman RG, Perlman DH, Serra M, et al. Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. Genome Biol. 2021;22(1):50.
11. Fulcher JM, Markillie LM, Mitchell HD, Williams SM, Engbrecht KM, Degnan DJ, et al. Parallel measurement of transcriptomes and proteomes from same single cells using nanodroplet splitting. Nat Commun. 2024;15(1):10614.
12. Gatto L, Aebersold R, Cox J, Demichev V, Derks J, Emmott E, et al. Initial recommendations for performing, benchmarking and reporting single-cell proteomics experiments. Nat Methods. 2023;20(3):375–86.
13. Zheng H, Vijg J, Fard AT, Mar JC. Measuring cell-to-cell expression variability in single-cell RNA-sequencing data: a comparative analysis and applications to B cell aging. Genome Biol. 2023;24(1):238.
14. Liu Q, Herring CA, Sheng Q, Ping J, Simmons AJ, Chen B, et al. Quantitative assessment of cell population diversity in single-cell landscapes. PLOS Biol. 2018;16(10):e2006687-e.
15. Bubis JA, Arrey TN, Damoc E, Delanghe B, Slovakova J, Sommer TM, et al. Challenging the Astral mass analyzer to quantify up to 5,300 proteins per single cell at unseen accuracy to uncover cellular heterogeneity. Nat Methods. 2025. https://doi.org/10.1038/s41592-024-02559-1.
16. Ai L, Binek A, Zhemkov V, Cho JH, Haghani A, Kreimer S, et al. Single cell proteomics reveals specific cellular subtypes in cardiomyocytes derived from human iPSCs and adult hearts. Mol Cell Proteomics. 2025. https://doi.org/10.1016/j.mcpro.2025.100910.
17. Lanz MC, Fuentes Valenzuela L, Elias JE, Skotheim JM. Cell size contributes to single-cell proteome variation. J Proteome Res. 2023;22(12):3773–9.
18. Rodríguez JM, Jorge I, Martinez-Val A, Barrero-Rodríguez R, Magni R, Núñez E, et al. iSanXoT: A standalone application for the integrative analysis of mass spectrometry-based quantitative proteomics data. Comput Struct Biotechnol J. 2024;23:452–9.
19. Trevisan-Herraz M, Bagwan N, García-Marqués F, Rodriguez JM, Jorge I, Ezkurdia I, et al. SanXoT: a modular and versatile package for the quantitative analysis of high-throughput proteomics experiments. Bioinformatics. 2019;35(9):1594–6.
20. García-Marqués F, Trevisan-Herraz M, Martínez-Martínez S, Camafeita E, Jorge I, Lopez JA, et al. A Novel Systems-Biology Algorithm for the Analysis of Coordinated Protein Responses Using Quantitative Proteomics. Mol Cell Proteomics. 2016;15(5):1740–60.
21. Sadek H, Olson EN. Toward the goal of human heart regeneration. Cell Stem Cell. 2020;26(1):7–16.
22. Bywater MJ, Burkhart DL, Straube J, Sabò A, Pendino V, Hudson JE, et al. Reactivation of Myc transcription in the mouse heart unlocks its proliferative capacity. Nat Commun. 2020;11(1):1827.
23. Ahuja P, Zhao P, Angelis E, Ruan H, Korge P, Olson A, et al. Myc controls transcriptional regulation of cardiac metabolism and mitochondrial biogenesis in response to pathological stress in mice. J Clin Investig. 2010;120(5):1494–505.
24 Villa del Campo C, Clavería C, Sierra R, Torres M. Cell competition promotes phenotypically silent cardiomyocyte replacement in the mammalian heart. Cell Rep. 2014;8(6):1741–51.
25. Villa del Campo C, Raiola M, Sierra R, Ms I, Peña-Peña J, Esteban I, et al. Myc overexpression improves recovery from myocardial infarction associated with cardiomyocyte hyperplasia in the mouse heart. Commun Biol. 2025;8(1):1069.
26. Kreimer S, Binek A, Chazarin B, Cho JH, Haghani A, Hutton A, et al. High-throughput single-cell proteomic analysis of organ-derived heterogeneous cell populations by nanoflow dual-trap single-column liquid chromatography. Anal Chem. 2023;95(24):9145–50.
27. Mund A, Coscia F, Kriston A, Hollandi R, Kovács F, Brunner A-D, et al. Deep Visual Proteomics defines single-cell identity and heterogeneity. Nat Biotechnol. 2022;40(8):1231–40.
28. Nordmann TM, Anderton H, Hasegawa A, Schweizer L, Zhang P, Stadler P-C, et al. Spatial proteomics identifies JAKi as treatment for a lethal skin disease. Nature. 2024;635(8040):1001–9.
29. Momenzadeh A, Jiang Y, Kreimer S, Teigen LE, Zepeda CS, Haghani A, et al. A Complete Workflow for High Throughput Human Single Skeletal Muscle Fiber Proteomics. J Am Soc Mass Spectrom. 2023;34(9):1858–67.
30. Vanderaa C, Gatto L. Replication of single-cell proteomics data reveals important computational challenges. Expert Rev Proteomics. 2021;18(10):835–43.
31 Vanderaa C, Gatto L. Scplainer: using linear models to understand mass spectrometry-based single-cell proteomics data. Genome Biol. 2025;26(1):237.

32. Clavería C, Giovinazzo G, Sierra R, Torres M. Myc-driven endogenous cell competition in the early mammalian embryo. Nature. 2013;500(7460):39–44.
33. Marín-Vicente C, Villa del Campo C, Calvo E, Rodríguez JM, Sierra R, Martín-Salamanca S, Torroja C, Végvári A, Zubarev R.A, Torres M, Vázquez J. Pro-regenerative fingerprints identified in a sub-population of adult mouse cardiomyocytes by integrative single-cell proteomics-Pilot dataset. ProteomeXchange-PRIDE. 2025. https://www.ebi.ac.uk/pride/archive/projects/PXD064499.
34. Vandenbulcke S, Vanderaa C, Crook O, Martens L, Clement L. msqrob2TMT: robust linear mixed models for inferring differential abundant proteins in labelled experiments with arbitrarily complex design. Mol Cell Proteom. 2025;24(7):101002.
35. Navarro P, Trevisan-Herraz M, Bonzon-Kulichenko E, Núñez E, Martínez-Acedo P, Pérez-Hernández D, et al. General Statistical Framework for Quantitative Proteomics by Stable Isotope Labeling. J Proteome Res. 2014;13(3):1234–47.
36. Marín-Vicente C, Villa del Campo C, Calvo E, Rodríguez JM, Sierra R, Martín-Salamanca S, Torroja C, Végvári A, Zubarev R.A, Torres M, Vázquez J. Pro-regenerative fingerprints identified in a sub-population of adult mouse cardiomyocytes by integrative single-cell proteomics-Analytical dataset. ProteomeXchange-PRIDE. 2025. https://www.ebi.ac.uk/pride/archive/projects/PXD064518.
37. Ackers-Johnson M, Li PY, Holmes AP, O'Brien S-M, Pavlovic D, Foo RS. A simplified, Langendorff-free method for concomitant isolation of viable cardiac myocytes and nonmyocytes from the adult mouse heart. Circ Res. 2016;119(8):909–20.
38. Ackers-Johnson M, Tan WLW, Foo R-Y. Following hearts, one cell at a time: recent applications of single-cell RNA sequencing to the understanding of heart disease. Nat Commun. 2018;9(1):4434.
39. Marín-Vicente C, Villa del Campo C, Calvo E, Rodríguez JM, Sierra R, Martín-Salamanca S, Torroja C, Végvári A, Zubarev R.A, Torres M, Vázquez J. Pro-regenerative fingerprints identified in a sub-population of adult mouse cardiomyocytes by integrative single-cell proteomics_FACS. ProteomeXchange-PRIDE. 2025. https://www.ebi.ac.uk/pride/archive/projects/PXD064501.
40. Marín-Vicente C, Villa del Campo C, Calvo E, Rodríguez JM, Sierra R, Martín-Salamanca S, Torroja C, Végvári A, Zubarev R.A, Torres M, Vázquez J. Pro-regenerative fingerprints identified in a sub-population of adult mouse cardiomyocytes by integrative single-cell proteomics_Complementary_data. Zenodo. 2026. https://doi.org/10.5281/zenodo.18379944.
41. Boulesteix AL, Strimmer K. Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. Brief Bioinform. 2006;8(1):32–44.
42. Suvarna K, Biswas D, Pai MGJ, Acharjee A, Bankar R, Palanivel V, et al. Proteomics and machine learning approaches reveal a set of prognostic markers for COVID-19 severity with drug repurposing potential. Front Physiol. 2021. https://doi.org/10.3389/fphys.2021.652799.
43. Xu S, Bai C, Chen Y, Yu L, Wu W, Hu K. Comparing univariate filtration preceding and succeeding PLS-DA analysis on the differential variables/metabolites identified from untargeted LC-MS metabolomics data. Anal Chim Acta. 2024;1287:342103-.
44. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44–57.
45. Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Res. 2022;50(W1):W216–21.
46. van Riggelen J, Yetil A, Felsher DW. MYC as a regulator of ribosome biogenesis and protein synthesis. Nat Rev Cancer. 2010;10(4):301–9.
47. Popay TM, Wang J, Adams CM, Howard GC, Codreanu SG, Sherrod SD, et al. MYC regulates ribosome biogenesis and mitochondrial gene expression programs through its interaction with host cell factor–1. eLife. 2021. https://doi.org/10.7554/eLife.60191.
48. Fujiwara Y, Naka Y, Sasalo M, Imahashi K, Yoshida Y. Identification of cell adhesion molecules for the maturation of engineered heart tissue. Eur Heart J. 2024. https://doi.org/10.1093/eurheartj/ehae666.3645.
49. Weber KT, Sun Y, Tyagi SC, Cleutjens JPM. Collagen network of the myocardium: function, structural remodeling and regulatory mechanisms. J Mol Cell Cardiol. 1994;26(3):279–92.
50. Rienks M, Papageorgiou A-P, Frangogiannis NG, Heymans S. Myocardial extracellular matrix. Circ Res. 2014;114(5):872–88.
51. Li F, Wang Y, Zeller KI, Potter JJ, Wonsey DR, O'Donnell KA, et al. Myc Stimulates Nuclearly Encoded Mitochondrial Genes and Mitochondrial Biogenesis. Mol Cell Biol. 2005;25(14):6225–34.
52. Maroli G, Braun T. The long and winding road of cardiomyocyte maturation. Cardiovasc Res. 2021;117(3):712–26.
53. Talman V, Teppo J, Pöhö P, Movahedi P, Vaikkinen A, Karhu ST, et al. Molecular atlas of postnatal mouse heart development. J Am Heart Assoc. 2018. https://doi.org/10.1161/JAHA.118.010378.
54. Boczonadi V, Giunta M, Lane M, Tulinius M, Schara U, Horvath R. Investigating the role of the physiological isoform switch of cytochrome c oxidase subunits in reversible mitochondrial disease. Int J Biochem Cell Biol. 2015;63:32–40.
55. García-Poyatos C, Arora P, Calvo E, Marques IJ, Kirschke N, Galardi-Castilla M, et al. Cox7a1 controls skeletal muscle physiology and heart regeneration through complex IV dimerization. Dev Cell. 2024;59(14):1824-41.e10.
56. Cluzeaut F, Maurer-Schultze B. Proliferation of cardiomyocytes and interstitial cells in the cardiac muscle of the mouse during pre- and postnatal development. Cell Tissue Kinet. 1986;19(3):267–74.
57. Chorghade S, Seimetz J, Emmons R, Yang J, Bresson SM, Lisio MD, et al. Poly(A) tail length regulates PABPC1 expression to tune translation in the heart. eLife. 2017. https://doi.org/10.7554/eLife.24139.
58. Porrello ER, Mahmoud AI, Simpson E, Hill JA, Richardson JA, Olson EN, et al. Transient Regenerative Potential of the Neonatal Mouse Heart. Science. 2011;331(6020):1078–80.
59. González-Rosa JM, Sharpe M, Field D, Soonpaa MH, Field LJ, Burns CE, et al. Myocardial Polyploidization Creates a Barrier to Heart Regeneration in Zebrafish. Dev Cell. 2018;44(4):433-46.e7.
60. Nishiyama C, Saito Y, Sakaguchi A, Kaneko M, Kiyonari H, Xu Y, et al. Prolonged Myocardial Regenerative Capacity in Neonatal Opossum. Circulation. 2022;146(2):125–39.

61 Yahya G, Menges P, Amponsah PS, Ngandiri DA, Schulz D, Wallek A, et al. Sublinear scaling of the cellular proteome with ploidy. Nat Commun. 2022;13(1):6182.
62. Claude K-L, Bureik D, Chatzitheodoridou D, Adarska P, Singh A, Schmoller KM. Transcription coordinates histone amounts and genome content. Nat Commun. 2021;12(1):4202-.
63. Marshall WF. Scaling of subcellular structures. Annu Rev Cell Dev Biol. 2020;36(1):219–36.
64. Chan Y-HM, Marshall WF. Scaling properties of cell and organelle size. Organogenesis. 2010;6(2):88–96.
65. Wesley CC, Mishra S, Levy DL. Organelle size scaling over embryonic development. WIREs Dev Biol. 2020. https://doi.org/10.1002/wdev.376.
66. Li X, Wu F, Günther S, Looso M, Kuenne C, Zhang T, et al. Inhibition of fatty acid oxidation enables heart regeneration in adult mice. Nat. 2023. https://doi.org/10.1038/s41586-023-06585-5.
67. Cardoso AC, Lam NT, Savla JJ, Nakada Y, Pereira AHM, Elnwasany A, et al. Mitochondrial substrate utilization regulates cardiomyocyte cell cycle progression. Nat Metab. 2020;2(2):167–78.
68. Fukuda R, Marín-Juez R, El-Sammak H, Beisaw A, Ramadass R, Kuenne C, et al. Stimulation of glycolysis promotes cardiomyocyte proliferation after injury in adult zebrafish. EMBO Rep. 2020. https://doi.org/10.15252/embr.201949752.
69. Hellicar J, Stevenson NL, Stephens DJ, Lowe M. Supply chain logistics – the role of the Golgi complex in extracellular matrix production and maintenance. J Cell Sci. 2022. https://doi.org/10.1242/jcs.258879.
70. Leone M, Engel FB. Pseudo-bipolar spindle formation and cell division in postnatal binucleated cardiomyocytes. J Mol Cell Cardiol. 2019;134:69–73.
71. Agah R, Frenkel PA, French BA, Michael LH, Overbeek PA, Schneider MD. Gene recombination in postmitotic cells. Targeted expression of Cre recombinase provokes cardiac-restricted, site-specific rearrangement in adult ventricular muscle in vivo. J Clin Invest. 1997;100(1):169–79.
72. Friedrich G, Soriano P. Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. Genes Dev. 1991;5(9):1513–23.
73. Bell RM, Mocanu MM, Yellon DM. Retrograde heart perfusion: the Langendorff technique of isolated heart perfusion. J Mol Cell Cardiol. 2011;50(6):940–50.
74. Marín-Vicente C, Calvo E, Rodríguez JM, Villa del Campo C, Sierra R, Végvári Á, et al. A Sample Preparation Procedure for Isobaric Labeling-Based Single-Cell Proteomics. 2024. p. 33–43.
75. Marín-Vicente C , Villa del Campo C, Calvo E, Rodríguez JM, Sierra R, Martín-Salamanca S, Torroja C, Végvári A, Zubarev R.A, Torres M, Vázquez J. Pro-regenerative fingerprints identified in a sub-population of adult mouse cardiomyocytes by integrative single-cell proteomics. Timeline analysis of bulk mice hearts. ProteomeXchange-PRIDE. 2025. https://www.ebi.ac.uk/pride/archive/projects/PXD064994.
76. Marín-Vicente C, Villa del Campo C, Calvo E, Rodríguez JM, Sierra R, Martín-Salamanca S, Torroja C, Végvári A, Zubarev R.A, Torres M, Vázquez J. Pro-regenerative fingerprints identified in a sub-population of adult mouse cardiomyocytes by integrative single-cell proteomics-Control and Myc overexpressing hearts. ProteomeXchange-PRIDE. 2026. https://www.ebi.ac.uk/pride/archive/projects/PXD073496.
77. Wiśniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. Nat Methods. 2009;6(5):359–62.
78. Furtwängler B, Üresin N, Motamedchaboki K, Huguet R, Lopez-Ferrer D, Zabrouskov V, et al. Real-Time Search-Assisted Acquisition on a Tribrid Mass Spectrometer Improves Coverage in Multiplexed Single-Cell Proteomics. Mol Cell Proteom. 2022;21(4):100219.
79. Perez-Riverol Y, Bandla C, Kundu DJ, Kamatchinathan S, Bai J, Hewapathirana S, et al. The PRIDE database at 20 years: 2025 update. Nucleic Acids Res. 2025;53(D1):D543-53.
80. Deutsch EW, Bandeira N, Perez-Riverol Y, Sharma V, Carver Jeremy J, Mendoza L, et al. The ProteomeXchange consortium in 2026: making proteomics data FAIR. Nucleic Acids Res. 2026;54(D1):D459–69.
81. Navarro P, Vázquez J. A refined method to calculate false discovery rates for peptide identification using decoy databases. J Proteome Res. 2009;8(4):1792–6.
82. Bonzon-Kulichenko E, Garcia-Marques F, Trevisan-Herraz M, Vazquez J. Revisiting peptide identification by high-accuracy mass spectrometry: problems associated with the use of narrow mass precursor windows. J Proteome Res. 2015;14(2):700–10.
83. Kong W, Hui HWH, Peng H, Goh WWB. Dealing with missing values in proteomics data. Proteomics. 2022. https://doi.org/10.1002/pmic.202200092.
84. Snel B. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. Nucleic Acids Res. 2000;28(18):3442–4.
85. Xia J, Psychogios N, Young N, Wishart DS. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. Nucleic Acids Res. 2009;37(Web Server):W652–60.
86 Pang Z, Chong J, Zhou G, de Lima Morais DA, Chang L, Barrette M, et al. MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. Nucleic Acids Res. 2021;49(W1):W388-96.
87. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184(13):3573-87.e29.
88. Anderson MJ. A new method for non-parametric multivariate analysis of variance. Austral Ecol. 2001;26(1):32–46.
89. Anderson MJ. Permutational Multivariate Analysis of Variance-PERMANOVA. Wiley StatsRef: Statistics Reference Online: Wiley; 2017. p. 1–15.
90. Smyth GK. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol. 2004;3(1):1–25.
91. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47-e.
92. Hoffman GE, Schadt EE. VariancePartition: interpreting drivers of variation in complex gene expression studies. BMC Bioinformatics. 2016;17(1):483.
93 Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, et al. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. Nucleic Acids Res. 2018;46(W1):W486-94.

94. Westerhuis JA, Hoefsloot HCJ, Smit S, Vis DJ, Smilde AK, van Velzen EJJ, et al. Assessment of PLSDA cross validation. Metabolomics. 2008;4(1):81–9.
95. Marín-Vicente C, Villa del Campo C, Calvo E, Rodríguez JM, Sierra R, Martín-Salamanca S, Torroja C, Végvári A, Zubarev R.A, Torres M, Vázquez J. Pro-regenerative fingerprints identified in a sub-population of adult mouse cardiomyocytes by integrative single-cell proteomics-Raw_data_1. Zenodo. 2026. https://doi.org/10.5281/zenodo.19484833.
96. Marín-Vicente C, Villa del Campo C, Calvo E, Rodríguez JM, Sierra R, Martín-Salamanca S, Torroja C, Végvári A, Zubarev R.A, Torres M, Vázquez J. Pro-regenerative fingerprints identified in a sub-population of adult mouse cardiomyocytes by integrative single-cell proteomics_Raw_data_2. Zenodo. 2026. https://doi.org/10.5281/zenodo.18379767.
97. Marín-Vicente C, Villa del Campo C, Calvo E, Rodríguez JM, Sierra R, Martín-Salamanca S, Torroja C, Végvári A, Zubarev R.A, Torres M, Vázquez J. Pro-regenerative fingerprints identified in a sub-population of adult mouse cardiomyocytes by integrative single-cell proteomics_Raw_data_3. Zenodo. 2026 https://doi.org/10.5281/zenodo.18375680.
98. Marín-Vicente C, Villa del Campo C, Calvo E, Rodríguez JM, Sierra R, Martín-Salamanca S, Torroja C, Végvári A, Zubarev R.A, Torres M, Vázquez J. Pro-regenerative fingerprints identified in a sub-population of adult mouse cardiomyocytes by integrative single-cell proteomics_Raw_data_4. Zenodo. 2026. https://doi.org/10.5281/zenodo.18375720.
99. Marín-Vicente C, Villa del Campo C, Calvo E, Rodríguez JM, Sierra R, Martín-Salamanca S, Torroja C, Végvári A, Zubarev R.A, Torres M, Vázquez J. Pro-regenerative fingerprints identified in a sub-population of adult mouse cardiomyocytes by integrative single-cell proteomics_iSanXoT. Zenodo. 2026. https://doi.org/10.5281/zenodo.19439667.
100. Marín-Vicente C, Villa del Campo C, Calvo E, Rodríguez JM, Sierra R, Martín-Salamanca S, Torroja C, Végvári A, Zubarev R.A, Torres M, Vázquez J. Pro-regenerative fingerprints identified in a sub-population of adult mouse cardiomyocytes by integrative single-cell proteomics. Github. 2025. https://cnic-proteomics.github.io/iSanXoT/.

Publisher's Note