Recent developments in RNA sequencing are now allowing us to study genome-wide gene expression differences in individual cells. Although still relatively in early days, knowledge obtained by single-cell RNA sequencing (scRNA seq) has already significantly improved our understanding in biology and disease and will undoubtedly continue to do so. However, as is true for many new technologies, scRNA seq still comes with limitations that confine the insights that can be gained from the acquired data. Both biology and current methodology impact the outcomes and restrict us in getting a complete and true view on genome-wide gene expression changes occurring at a single-cell level. Although further improvements will surely resolve at least a part of these limitations, as yet, we should be mindful of how to appropriately mine the data to advance our knowledge of molecular mechanisms relevant for biology.

Magnifying Our View on Cardiac Gene Expression

Never before were we, as scientists, able to get such a detailed view on the molecular changes driving biology. Where conventional bulk sequencing inherently dilutes out lower signals and loses any information on cellular origin, scRNA seq provides detailed transcriptome-wide profiles of each individual cell that is being analyzed. This innovation already has allowed for the identification of novel stem cell populations and rare cell types. Although originally applied to in vitro, more recently, it was used to examine cells from digested mammalian tissues, like skeletal muscle, pancreas, and the heart.

The first reports on cardiac scRNA seq focused on gene expression profiles in cardiac cells derived from human-induced pluripotent stem cells and individual cells of the embryonic heart. Here, analysis at single-cell resolution revealed that current differentiation protocols for cardiac endothelial cells produce 4 different subpopulations of endothelial cells. Additionally, these studies showed us that distinct gene sets are expressed in cardiomyocytes from specific anatomic sites. Something else we learned that is maybe a bit less obvious, yet important to realize, is that there is no distinct gene expression profile in cells representing a certain phase of development. Instead of going through distinct phases of gene expression, it is the continuous gradient of gene expression that defines the different stages of cardiomyocyte differentiation. What separates a more mature cardiomyocyte from a nascent one is not the specific set of genes it expresses but rather the combined relative expression level of these genes.

More recently, scRNA seq was also successfully performed on the adult heart, where all main cardiac cell types could be detected under both healthy and diseased conditions. Gradients of gene expression were also found here but then present within the different subpopulations of a specific cell type. These data revealed that separate subclusters of adult cardiomyocytes showed a high level of heterogeneity in expression of well-established cardiomyocyte markers, such as Actc1 (actin, alpha, cardiac muscle 1) and Myh6 (α-myosin heavy chain). Similar gradients were seen in other noncardiomyocyte cell types of the heart. These observations imply that based on gene expression differences, a cell type, even under basal conditions, can consist of several different subpopulations that are functionally different. Also, when making use of cardiomyocyte-specific promoters, such as the Myh6 promoter, one should consider that the activity might vary from one cell to another.

Interestingly, in mining the scRNA seq data, it also became clear that there is gene expression contamination across the different cell types. Marker gene expression is often not purely restricted to a certain cell type, just strongly enriched in cells it is suppose to mark. For example, some cells that are clearly cardiomyocytes based on their transcriptional profile also express genes that are said to be endothelial or fibroblast specific (eg, Tie1 [tyrosine kinase with immunoglobulin-like and EGF-like domains 1] or Pecam1 [platelet endothelial cell adhesion molecule] for endothelial cells and Col1a1 [collagen, type I, alpha 1] and Pdgfra [platelet-derived growth factor receptor A] for fibroblasts), albeit at substantially lower levels. Above findings challenge our binary thinking about expression of markers in cell types. Just as with the maturation stages during development, cell types do not seem to be determined by the presence or absence of marker expression but rather by the level of expression of genes relevant for determining cell identity.
Room for Improvements

Although scRNA seq in the heart yields numerous possibilities to increase our understanding and place nuances in cardiac biology and pathology, currently, there are some consequential limitations and methodological issues that can result in misinterpretation of the data and could distort scientific conclusions (Figure).

Currently, the biggest issue with scRNA seq in the heart is the limited coverage, which is predicted to pick up only 5% to 20% of the full transcriptome per cell.11,12 This means that we only obtain information about the most abundant genes. In addition, a large portion of the sequencing reads is mapped to highly expressed mitochondrial genes and can, therefore, not serve to provide information on nuclear gene expression.5 In mitochondria-rich cardiomyocytes, ≤70% to 80% of all sequencing reads are soaked up by mitochondrial genes. A quick calculation then suggests that only 1% to 6% of all reads originate from genomic genes, severely restricting the gene expression information that can be gained per cardiomyocyte. As a result, most genes detected in cardiomyocytes are highly expressed cardiomyocyte markers, whereas lower expressed genes will only sporadically be picked up. The low coverage renders it particularly hard to detect rare subtypes within the cell populations if they are characterized by low expressing genes. This becomes especially relevant when one is trying to study rare cardiac progenitor cells or attempting to identify the exceptional de-differentiating and proliferating cardiomyocyte. Only when a high number of cells are being sequencing, the chance for detecting these rare cellular subtypes will increase. Current studies that sequenced the adult heart either did not sequence cardiomyocytes or only a limited number.5,10,13 A higher number of cells is required to obtain a reliable and full representation of the transcriptomic landscape across the adult cardiomyocyte population.

As such, although these limitations exist, it is important to realize that the absence of evidence for rare cell populations in the heart is not evidence of absence and can certainly to date not be excluded by using scRNA seq.

Because transcriptomes of single cells are only measured at one time point and not over time, it is difficult to determine whether transcriptional heterogeneity between cells is really because of biology or whether it arises because of stochastic gene expression changes over time. As such, it is possible that based on gene expression similarities, certain cells are clustered as a separate subpopulation of cells, whereas the overlap in gene expression profile is merely because of temporal synchrony.

Various sorting strategies are currently used to separate single cells into individual wells or droplets. However, all these strategies have physical constraints on the cell size that they sort. Commercially available single-cell sorting platforms like Fluidigm C1 and Chromium can currently only sort cells that are ≤25 to 50 µm in diameter. This is considerably smaller than adult mammalian cardiomyocytes, which can be ≈125 µm along the longitudinal axis.14 Sequencing adult cardiomyocytes on current commercial systems might prove difficult because the cells are simply too big to effectively pass through the sorting system. Although we recently were able to sort adult cardiomyocytes,5 this approach is relatively expensive and proved biased toward the larger cardiomyocytes. Although it is thought that roughly 30% of the cells in the adult heart are cardiomyocytes, 75% of the cells that we sequenced turned out to be cardiomyocytes and captured less of the smaller cardiac cell types as a trade-off. Because of these different cellular biases in sorting strategies, the scRNA seq technology is currently unable to reliably capture the relative abundance of all adult cardiac cell types in an unbiased fashion.

In time, more improvements will be made to existing scRNA seq technology, and new, presumably superior methods will be developed. Although this is great for the depth and reliability of the data this technology provides, these advancements might actually pose a problem by itself because they make it increasingly difficult and potentially even impossible to compare and combine the different datasets that are now being generated. Differences between experiments in transcriptomic coverage per cell make it complex to compare the different data sets. In addition, new deep-sequencing techniques like nanopore sequencing generate whole new types of datasets, which could be challenging to combine and compare with current Illumina-based sequencing data.

Figure. Schematic overview of the single-cell sequencing technique on cardiac tissue. The red boxes depict potential technical limitations that can hinder scientific interpretation of the data sets. FACS indicates fluorescence-activated cell sorting; t-SNE, t-distributed stochastic neighbor embedding.
Are You Part of the Solution or Just Here to Complain?

Despite the indicated points of caution, scRNA seq has an unprecedented potential to advance our understanding of cardiac biology and pathology. For this reason, we should put a lot of effort into capitalizing this technology.

One potential way to surpass the relatively low resolution in gene expression in adult cardiomyocytes would be to exclude the high number of mitochondrial reads. Because there are only 37 mitochondrial genes, it should be possible to eliminate mitochondrial genes from the sequencing library. This should significantly free up reagents for genomic transcripts and increase the depth of information gained about genomic gene expression. Furthermore, further improvements in commercially available single-cell sorting systems will potentially enable high-throughput sorting of large cells, like cardiomyocytes. This will increase the number of cardiomyocytes that can be sequence form adult mammalian hearts.

As with all knowledge obtained with new techniques, findings should be validated using more established experimental techniques. For example, true subtype-specific differences between cells or disease-induced expressional changes should be validated using in situ hybridization or immunohistochemistry. This should also prevent misinterpretation of data because of dissociation-induced artifacts on gene expression profiles. Reproduction of scRNA seq experiments on independent sample sets will further increase the reliability of the data sets.

No doubt near future adjustments will lead to further optimization of the scRNA seq approach for the heart. However, even with the current limitations, this technology enabled us to learn more about cardiac development, transcriptional heterogeneity within cell types, and cell type-specific gene expression changes during cardiac disease. It will be exciting to see how advances in scRNA seq techniques will increase transcriptome information at single-cell resolution and how it will continue to improve our knowledge about heart biology and disease.

Sources of Funding

This work was supported by the Leducq Foundation (14CVD04 to E. van Rooij) and the European Research Council (ERC) under the European Union’s Seventh Framework Programme (ERC grant agreement CoG 615708 MICARUS to E. van Rooij).

Disclosures

None.

References

1. Grün D, Lyubimova A, Kester L, Wiebrands K, Basak O, Sasaki N, Clevers H, van Oudenaarden A. Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature*. 2015;525:251–255. doi: 10.1038/nature14966
2. Grün D, Muraro MJ, Boisset JC, Wiebrands K, Lyubimova A, Dharmadhikari G, van den Born M, van Es J, Jansen E, Clevers H, de Koning EJP, van Oudenaarden A. De novo prediction of stem cell identity using single-cell transcriptome data. *Cell Stem Cell*. 2016;19:266–277. doi: 10.1016/j.stem.2016.05.010
3. van der Brink SC, Sage F, Vertesy A, Spanjaard B, Peterson-Maduro J, Barons CS, Robin C, van Oudenaarden A. Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat Methods*. 2017;14:955–936. doi: 10.1038/nmeth.4437
4. Muraro MJ, Dharmadhikari G, Grün D, Groen N, Dielen T, Jansen E, van Gurp L, Engelse MA, Carlotti F, de Koning EJ, van Oudenaarden A. A single-cell transcriptome atlas of the human pancreas. *Cell Syst*. 2016;3:385.e3–394.e3. doi: 10.1016/j.cels.2016.09.002
5. Gldaka MM, Molenaar B, de Ruiter H, van der Elst S, Tsui H, Versteeg D, Lacraz GPA, Huibers MMH, van Oudenaarden A, van Rooij E. Single-cell sequencing of the healthy and diseased heart reveals cytoskeleton-associated protein 4 as a new modulator of fibroblasts activation. *Circulation*. 2018;138:166–180. doi: 10.1161/CIRCULATIONAHA.117.030742
6. Paik DT, Tian L, Lee J, Sayed N, Chen IY, Rhee S, Rhee JW, Kim Y, Wirka RC, Buikema JW, Wu SM, Red-Horse K, Quertermous T, Wu JC. Large-scale single-cell RNA-Seq reveals molecular signatures of heterogeneous populations of human induced pluripotent stem cell-derived endothelial cells. *Circ Res*. 2018;123:443–450. doi: 10.1161/CIRCRESAHA.118.312913
7. Li G, Xu A, Sim S, Priest JR, Tian X, Khan T, Quertermous T, Zhou B, Tsaö PS, Quake SR, Wu SM. Transcriptomic profiling maps anatomically patterned subpopulations among single embryonic cardiac cells. *Dev Cell*. 2016;39:491–507. doi: 10.1016/j.devcel.2016.10.014
8. DeLaughter DM, Bick AG, Wakimoto H, McKean D, Gontham JM, Kathiriya IS, Hinson JT, Homey J, Gray J, Pu W, Bruneau BG, Seidman JD, Seidman CE. Single-cell resolution of temporal gene expression during heart development. *Dev Cell*. 2016;39:480–490. doi: 10.1016/j.devcel.2016.10.001
9. Xiao Y, Hill MC, Zhang M, Martin TJ, Morikawa Y, Wang S, Moise AR, Wythe JD, Martin JF. Hippo signaling plays an essential role in cell state transitions during cardiac fibroblast development. *Dev Cell*. 2018;45:153.e6–169.e6. doi: 10.1016/j.devcel.2018.03.019
10. Skelly DA, Squiers GT, McLellan MA, Bolisetty MT, Robson P, Rosenthal NA, Pinto AR. Single-cell transcriptional profiling reveals cellular diversity and intercommunication in the mouse heart. *Cell Rep*. 2018;22:600–610. doi: 10.1016/j.celrep.2017.12.072
11. Grün D, Kester L, van Oudenaarden A. Validation of noise models for single-cell transcriptomics. *Nat Methods*. 2014;11:637–640. doi: 10.1038/nmeth.2930
12. Hashimshony T, Senderovich N, Avital G, Kloechlender A, de Leeuw Y, Anavy L, Gennert D, Li S, Livak KJ, Rozenblatt-Rosen O, Regev K, Yanai I. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol*. 2016;17:77. doi: 10.1186/s13059-016-0938-8
13. Schaffer S, Viswanathan S, Widjaja AA, et al. HH-11 is a crucial determinant of cardiovascular fibrosis. *Nature*. 2017;552:110–115. doi: 10.1038/nature24676
14. Sorenson AL, Tepper D, Sonnenblick EH, Robinson TF, Capasso JM. Size and shape of enzymatically isolated ventricular myocytes from rats and cardiomyopathic hamsters. *Cardiovasc Res*. 1985;19:793–799.

Key Words: gene expression • genome • heart • myocytes, cardiac • read count • single-cell sequencing