Molecular Aspects of Medicine xxx (2017) 1–4

Contents lists available at ScienceDirect

Molecular Aspects of Medicine

journal homepage: www.elsevier.com/locate/mam

Editorial

The secrets of the cell

Keywords:
Cell heterogeneity
Cell types
Single-cell analysis
Single-cell biology

2017 marks the 40 year's anniversary of the development of nucleic acid sequencing independently by Fred Sanger and Walter Gilbert (Sanger et al., 1977; Maxam and Gilbert, 1977), who shared the Nobel Prize in Chemistry in 1980. It is also 34 years since Kary Mullis conceived the polymerase chain reaction (PCR), for which he shared the Nobel Prize in Chemistry in 1993. PCR was then put into practice in 1985 by a team of Cetus scientists (Saiki et al., 1985). Just three years later came the first publication demonstrating the sensitivity of PCR by analyzing DNA from a single somatic and a single sperm cell (Li et al., 1988). Using classical PCR analyzing DNA, no cell to cell variability was detected. This had to await the development of real-time quantitative PCR (qPCR) by Russ Higuchi in 1993 (Higuchi et al., 1993). Several single-cell gene expression studies appeared in the late 90-ies (Freeman et al., 1999) and in 2005 came the first comprehensive single-cell reverse transcription (RT) qPCR profiling study (Bengtsson et al., 2005). It was performed on primary mouse beta-cells and revealed unexpected heterogeneity in the amount of transcripts across the cells, with majority of the cells containing just a few of each targeted transcript, while some few cells harboring large numbers (Fig. 1A). This variation of transcripts across similar cells was modelled with a normal distribution in logarithmic scale (Fig. 1B). Next generation sequencing methods were developed that made whole transcriptome profiling possible and were coined Method of the Year by the journal Nature Methods in 2013 (Method of the Year, 2013). Today, we see rapid development of high throughput single-cell RNA sequencing platforms that allow the profiling of hundreds of thousands of cells at affordable cost. In parallel, methods are being developed for multimodal profiling allowing the simultaneous quantification of RNA, protein, and DNA in single cells (Ståhlberg et al., 2012), as well as characterizing epigenetic modifications (Clark et al., 2016). The single cell era we are approaching is expected to lead to new insights into biology, novel discoveries and possibly even challenge some dogmas. Particularly exciting will be the new possibilities to characterize cell types and study their differentiation, proliferation and function. The tens of trillions (10^{13}) of cells in a human body are often said to be made up of some 210 cell types subdivided into 20 categories assembled in 1989 based primarily on function (Alberts et al., 1989). A more recent classification suggests there are 411 cell types (Vickaryous and Hall, 2006). To address this further the Human Cell Atlas initiative was launched 2016 with the aim to create comprehensive reference maps of all human cells (Human Cell Atlas). However, a precise and unambiguous definition of cell type is notoriously difficult. Environmental conditions, external stimuli, number and nature of neighboring cells, signals from remote cells through hormones, exosomes and other signaling substances, access to nutrients, oxygen and other vital substances, removal of waste products, phase of cell cycle, accumulated somatic mutations, integrated viruses, transposons, epigenetic alterations, chromosomal rearrangements, copy number variations, epigenetic modifications, and perhaps even age and generation will affect a cell's molecular activities. Some may lead to virtually irreversible differentiated states, while other may lead to reversible or even temporal changes only. Single-cell profiling is expected to shed light on these processes, perhaps by identifying cell type-specific expression networks that will contribute to establishing a definition of cell type and defining the molecular events that make a change virtually irreversible.

Decomposing a sample into cells that are profiled individually we expect to see a lot of variation. Based on our current knowledge there are at least four possible sources to variation: 1) **Cell type**. Although the traditional definition of cell type is outdated and ambiguous, generally we think of different cell types as being phenotypically spontaneously non-interconvertible. Most biological samples are composed of different types of cells that show different molecular expressions and functions. 2) **Microenvironment**. A cell's nearest surrounding and interaction with neighboring cells affect its expression as well as stimuli such as hormones, exosomes, and other signaling substances including specific ions, peptides, amino acids, metabolites and many other low molecular weight substances such as exogenous drugs. Also viruses, acute infections as well as many of the chronic infections and modifications acquired throughout life, and bacteria affect cells' expressions. Even cells in culture show remarkable differences in expression depending on their location and surrounding. For example, cells in less confluent areas may have better access to nutrients and easier to dispose of waste products than densely packed cells, which may impact on their expression. The response to changes in the microenvironment can be very fast and we expect effects of removing the cells from their natural environment in the live tissue, the collection process, handling and sorting of cells to influence the expression of sensitive genes. For traditional samples this is addressed by the SPIDIA consortium (Spidia) that generates data and drafts guidelines for the preanalytical process in molecular diagnostics. For example, sampling of blood in EDTA tubes alters transcript levels of some genes in leukocytes 20-fold

http://dx.doi.org/10.1016/j.mam.2017.08.004

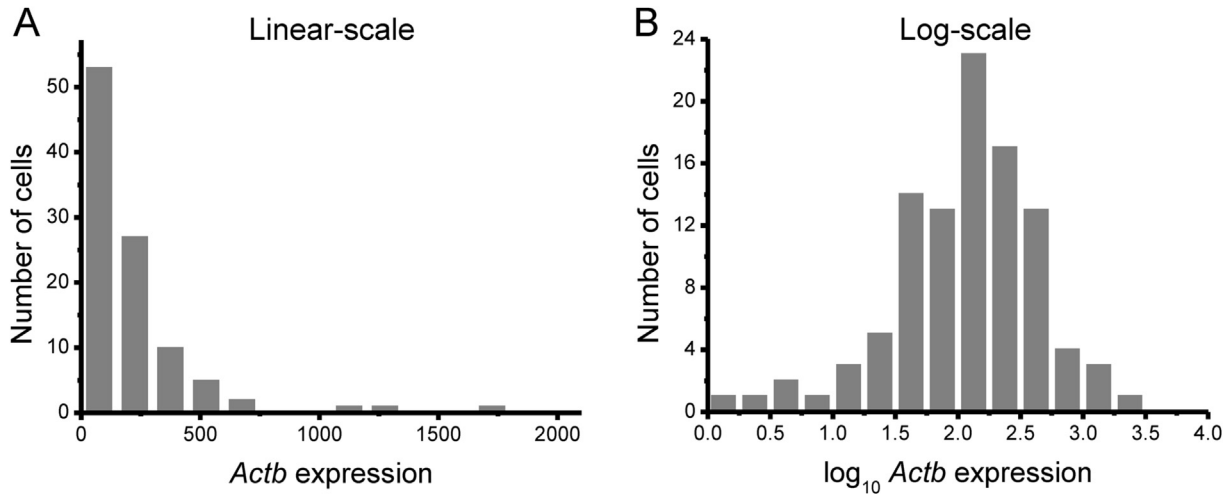

Fig. 1. Expression of mRNA molecules in individual cells. Histograms showing the expression of *Actb* in 96 individual primary mouse beta-cells in (A) linear scale and (B) logarithmic scale (Bengtsson et al., 2005).

(Pahl and Brune, 2002; Pazzagli et al., 2013). 3) **Cell state dependence**. In tissues and organs there is a need to sustain the cell mass by cell division, growth and apoptosis. Undifferentiated stem cells are often quiescent and divide upon specific signaling, while progenitor cells are highly proliferative and develop into specialized and differentiated cells that are senescent. Although majority of genes are expressed independently of cell cycle, many have been found cell cycle dependent (Ly et al., 2014). 4) **Temporal variations.** Even seemingly identical cells from a cell line show large variation in transcript levels that has been attributed to burst kinetics of gene expression (Chubb et al., 2006; Raj et al., 2006). Currently we are only aware of one exception to this lognormal distribution of transcripts: it is not seen across amphibian oocytes, which lack mRNA metabolism (unpublished data). Table 1 shows various factors that have been reported to affect cells expressions in seemingly homogenous cell populations.

Due to the temporal variation of genes' and proteins' expressions it is usually not possible to tell two cell types apart from the expression level of a single marker. Rather, to distinguish different cell types correlation between genes' expressions should be exploited. Genes are not expressed independently of each other, rather genes in the same expression pathway or network tend to be expressed in the same cell at the same time, presumably because of having correlated expression bursts. Those genes can be identified using multivariate methods, which clusters the cells based on their expression profiles (Bergkvist et al., 2010). In multivariate analyses only genes responsive to the conditions studied should be included. Genes, which expression is affected by the handling of the cells must be excluded, and also non-responsive genes shall be removed from analysis as they only contribute with noise. This can be challenging in high-throughput global single-cell profiling studies as it may be hard to identify the genes whose transcription is affected by

Table 1
Overview of various factors causing cell heterogeneity.

Cell type	Factor causing cell heterogeneity	Reference
Human breast cancer cell lines	Anoikis resistance, hypoxia, sphere forming assays	Akrap et al., 2016
Escherichia coli	Antibiotic resistance	Baltekin et al., 2017
Mouse beta-cells	Glucose	Bengtsson et al., 2005
Mouse embryonic stem cells	Cell cycle	Buettner et al., 2015
Human T-cells	Pathogens	Bushkin et al., 2015
Dictyostelium	Transcriptional bursting	Chubb et al., 2006
Human B-cells	Ageing	de Bourcy et al., 2017
Mouse preimplantation embryos	Allele specific expression	Deng et al., 2014
Human cell lines	Cell size	Dolatabadi et al., 2017
chronic myeloid leukemia cancer stem cells	Tumor development, mutations, environment	Giustacchini et al., 2017
Human circulating tumor cells	Tumor development, mutations, environment	Gorges et al., 2016
Human pre-implantation embryos	Early human development	Guo et al., 2014
Caenorhabditis elegans	Wnt signaling	Ji et al., 2013
Mouse brain cells	Transcription start sites	Karlsson et al., 2017
Mouse oligodendrocyte	Differentiation	Marques et al., 2016
Human neuroblastoma cell lines	RNA editing	Mellis et al., 2017
Induced pluripotent human embryonic stem cells	Dedifferentiation	Narsinh et al., 2011
Chinese hamster ovary cell lines	Transcriptional bursting	Raj et al., 2006
Mouse embryonic stem cells	miRNA regulation	Schmiedel et al., 2015
Human derived melanoma cells	Drug treatment	Shaffer et al., 2017
Xenopus laevis oocyte	Cell polarity	Sidova et al., 2015
Caenorhabditis elegans	LIN-3/EGF morphogen gradient	van Zon et al., 2015
Human sperm cells	Recombination and mutation rate	Wang et al., 2012
Human lymphoblastoid cell lines	Single-nucleotide polymorphisms	Wills et al., 2013
Human osteosarcoma cell line and mouse neurons	Translation bursting	Wu et al., 2016

Please cite this article in press as: Kubista, M., et al., The secrets of the cell, Molecular Aspects of Medicine (2017), http://dx.doi.org/10.1016/j.mam.2017.08.004