Genome-wide identification of genes essential for podocyte cytoskeletons based on single-cell RNA sequencing

see commentary on page 1038

Yuqiu Lu[1,2], Yuting Ye[1,2], Wenduona Bao[1], Qianqian Yang[1], Jinquan Wang[1], Zhihong Liu[1] and Shaolin Shi[1]

[1]National Clinical Research Center for Kidney Diseases, Jinling Hospital, Nanjing University School of Medicine, Nanjing, Jiangsu 210002, China

Gene expression differs substantially among individual cells of the same type. We speculate that genes that are expressed in all but a portion of cells of a given cell type would be likely essential and required for either the cell survival (housekeeping) or for the cell type's unique structure and function, enabling the organism to survive. Here, we performed RNA-seq of 20 mouse podocytes using the Fluidigm C1 system and identified 335 genes that were expressed in all of them. Among them, 239 genes were also expressed in mesangial and endothelial cells and were involved in energy metabolism, protein synthesis, etc., as housekeeping genes. In contrast, 92 genes were preferentially expressed in podocytes (over five-fold versus expression in mesangial and endothelial cells) and are, therefore, the essential candidate genes specific for podocytes. Assessments by bioinformatics, conserved expression in human podocytes, and association with injury/disease all support the essentiality of these genes for podocytes. Factually, 27 of the 92 genes are already known to be essential for podocyte structure and function. Thirty-seven novel genes were functionally analyzed by siRNA silencing, and we found that a deficiency of 30 genes led to either cytoskeletal injury (FGFR1, AOX1, AIF1L, HAUS8, RAB3B, LPIN2, GOLIM4, CERS6, ARHGEF18, ARPC1A, SRGAP1, ITGB5, ILDR2, MPP5, TSC22D1, DNAJC11, SEPT10, MOCS2, FNBP1L, and TMOD3) or significant downregulation of CD2AP and synaptopodin (IFT80, MYOM2, ANXA4, CYB5R4, GPC1, ZNF277, NSF, ITGAV, CRYAB, and MTSS1). Thus, the list of genes essential for podocyte cytoskeletons is expanded by single-cell RNA sequencing. It appears that podocyte-specific essential genes are mainly associated with podocyte cytoskeletons.

Kidney International (2017) **92,** 1119–1129; http://dx.doi.org/10.1016/j.kint.2017.04.022

KEYWORDS: cytoskeleton; gene expression; podocyte

Correspondence: *Shaolin Shi, National Clinical Research Center of Kidney Disease, Jinling Hospital, Nanjing University School of Medicine, 305 East Zhongshan Road, Nanjing, Jiangsu 210002, China. E-mail: shaolinshi1001@yahoo.com*

[2]These authors contributed equally to the study.

Received 20 January 2017; revised 6 April 2017; accepted 13 April 2017; published online 12 July 2017

The bodies of higher animals comprise several hundred distinct cell types that are specialized for diverse functions and are required for the organism to survive and reproduce.[1] Elucidation of the molecular mechanisms that underlie unique functions of each cell type is important and relies on the identification of genes that are involved in the unique functions of specific cell types. However, currently, a high-throughput method for identifying genes that are involved in the unique structures and functions of specific cell types is lacking but is highly desired.

Single-cell RNA sequencing is emerging as a powerful tool for molecular biology research. Moreover, gene expression analyses of single cells can reveal molecular processes more precisely than those of cell populations and are extremely powerful tools for identifying subpopulations of specific cell types, for distinguishing among cell lineages, and for determining differential responses of individual cells to stimuli.[2,3] Surprisingly, single-cell RNA-seq has revealed that gene expression among individual cells of the same type substantially varies, with correlation coefficients being as low as 0.1.[2,4,5] Thus, we speculate that genes expressed in some cells, but not in other cells of the same type, are more likely dispensable for cell survival and specific structures and functions, which are essential for survival. In contrast, genes expressed in all cells of a cell type are more likely essential and are either required for survival (housekeeping genes) or for unique structures and functions, which are required for survival. Differentiating between the two categories of essential genes is possible by comparing gene expression profiles with those of other cell types because genes shared by other cell types are likely housekeeping genes, whereas the remaining ones would be those specifically essential for the cell type of interest.

Glomerular podocytes are highly specialized cells and are characterized by foot processes through which adjacent podocytes interdigitate and form thin slits for filtration. An increasing number of genes have been identified as essential for unique structures and functions of podocytes, including those encoding nephrin, podocin, and cluster of differentiation 2-associated protein (CD2AP).[6–9] Dysfunctions of these genes disrupts podocyte structure, leading to proteinuria and various glomerular diseases.[10–13] These genes were mainly identified by genetic, biochemical, and gene expression profiling approaches. However, these approaches are not efficient, even

in the case of microarray gene expression profiling. Microarray analyses can identify large numbers of genes with down-regulated expression, but few of the genes identified using this approach are demonstrated to be essential for podocyte structure and function. Hence, podocyte-specific essential genes remain largely unknown, and with fewer than 100 identified candidates,[14] the molecular basis of the unique podocyte structures and functions and the mechanisms that underlie podocyte injury remain poorly understood. Therefore, a high-throughput approach for identifying podocyte-specific essential genes is strongly desired.

We recently performed single-cell RNA-seq analyses of mouse mesangial cells and identified genes that were expressed in all single mesangial cells and thus were considered essential for mesangial cells.[15] We observed high enrichment of genes associated with endothelial cells, and these data were consistent with the categorization of mesangial cells as a type of pericyte.[16] These data illustrate the feasibility of single-cell RNA-seq–based approaches for identifying essential genes for a given cell type.In this study, we extended single-cell RNA-seq analysis to mouse podocytes and identified genes that were expressed in all 20 sequenced single podocytes. Subsequently, we compared the expression levels of these genes in podocytes with those in non-podocyte glomerular cells and identified 92 genes that are preferentially expressed in and specifically essential for podocytes. We validated the essentiality of these genes for podocytes and identified 30 novel genes that could be essential for podocyte cytoskeleton assembly. Thus, this study demonstrated the feasibility of identifying cell type-specific essential genes using single-cell RNA-seq analyses.

RESULTS

Design of the approach

We designed an approach (Figure 1) for the genome-wide identification of genes that are essential for podocyte structure and function (hereafter referred to as "podocyte-specific essential gene candidates") on the basis of single-cell RNA-seq analyses. First, 20 individual podocytes were subjected to RNA-seq and genes expressed in all cells were identified. Next, we compared gene expression levels in podocytes with those in glomerular mesangial and endothelial cells, which were simultaneously generated, and identified common genes (potential housekeeping genes) and podocyte-specific essential gene candidates. Finally, we assessed the reliability of these gene candidates through bioinformatics analyses, conserved expression in human podocytes, associations of expression levels with podocyte injury or disease, enrichment of genes known to be essential for podocytes, and functional assays using small, interfering RNA (siRNA) knockdown in podocytes.

Single podocyte isolation, cDNA synthesis, and RNA sequencing

Mouse glomeruli were isolated and dissociated into single cells and were then loaded onto the Fluidigm C1 Single-Cell Auto Prep System for single-cell separation and cDNA synthesis, according to recently described methods.[15] Subsequent quantitative polymerase chain reaction (qPCR) analyses of the podocyte markers podocin, synaptopodin (SYNPO), and Wt1 were performed to distinguish podocyte cDNA samples from those of other glomerular cell types. Sequencing libraries were constructed for 20 single podocytes from their cDNA samples and were then ultradeep sequenced. Averaged total reads, reads mapped, and numbers of genes detected are summarized in Table 1, and the details can be found in Supplementary Tables S1 to S3 and Supplementary Figure S1. The quality of single-podocyte RNA-seq is also demonstrated by the distributions of reads mapped to several genes, including the known podocyte marker genes Wt1 and podocin (Nphs2; Figure 2a). The 20 single-podocyte RNA-seq data have been deposited in the Gene Expression Omnibus (GSE88814). In addition, we simultaneously analyzed 14 single mesangial cells using the current methods and deposited the resulting RNA-seq data in the Gene Expression Omnibus with the accession number GSE92650.

Wide heterogeneity of gene expression among individual podocytes

The correlation coefficient (Pearson) for gene expression between any two podocytes was $R = 0.27 \pm 0.035$ (mean $\pm$ SD; Supplementary Table S4), which is similar to that of other cell types.[15,17,18] This finding demonstrates that as with other cell types, individual podocytes have substantially different gene expression profiles. Large differences in gene expression levels among individual podocytes were directly identified from reads per kilobase per millions (RPKMs) of the first 50 genes in the present 20 podocytes (Dataset 1 in the Supplementary Material), and most expressed genes had widely varying RPKM values across the 20 podocytes, ranging from 0 to 100s and even 1000s (e.g., ENSMUSG00000030082). To validate these observations of heterogeneity in RNA-seq data, we analyzed 30 genes in two podocyte cDNA samples using quantitative polymerase chain reaction and compared their expression levels with corresponding RPKMs from sequencing analyses of the two podocytes. The two sets of expression data were highly consistent and had a correlation coefficient of 0.81 (Supplementary Figure S2). To further demonstrate the reliability of the RNA-seq data, we compared RPKMs from a single podocyte with bulk podocyte RPKMs[19] and calculated a correlation coefficient of 0.475 (Figure 2b). Subsequently, we compared mean RPKMs of the 20 podocytes with those of bulk podocytes and calculated an elevated correlation coefficient of 0.767 (Figure 2c). This finding indicates that single podocytes recapitulated the activities of bulk podocytes in RNA-seq analyses, thus supporting the reliability of our single-cell data. It is expected that a higher correlation is achieved if the RNA-seq data regarding more single podocytes are combined.

Identification of 335 genes expressed in all 20 podocytes

Because RPKM values between 0.1 and 1.0 are commonly used in the literature,[20,21] we used 0.5 RPKM value as the