OPEN

Lineage-tracing methods and the kidney

Benjamin D. Humphreys[1,2,3] and Derek P. DiRocco[1,2]

[1]Renal Division, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA; [2]Harvard Medical School, Boston, Massachusetts, USA and [3]Harvard Stem Cell Institute, Cambridge, Massachusetts, USA

The kidney is a complex organ with over 30 different cell types, and understanding the lineage relationships between these cells is challenging. During nephrogenesis, a central question is how the coordinated morphogenesis, growth, and differentiation of distinct cell types leads to development of a functional organ. In mature kidney, understanding cell division and fate during injury, regeneration and aging are critical topics for understanding disease. Genetic lineage tracing offers a powerful tool to decipher cellular hierarchies in both development and disease because it allows the progeny of a single cell, or group of cells, to be tracked unambiguously. Recent advances in this field include the use of inducible recombinases, multicolor reporters, and mosaic analysis. In this review, we discuss lineage-tracing methods focusing on the mouse model system and consider the impact of these methods on our understanding of kidney biology and prospects for future application.

Kidney International (2014) **86,** 481–488; doi:10.1038/ki.2013.368; published online 2 October 2013

KEYWORDS: kidney development; kidney disease; lineage tracing; mouse genetics; recombination

Correspondence: *Benjamin D. Humphreys, Harvard Institutes of Medicine, Room 554, 4 Blackfan Circle, Boston, Massachusetts 02115, USA. E-mail: bhumphreys@partners.org*

Received 23 April 2013; revised 16 May 2013; accepted 23 May 2013; published online 2 October 2013

In a lineage-tracing experiment, a cell is identified by expression of a reporter gene. If that cell divides, expression of the reporter is passed on to all progeny, and migration of these marked cells can be easily measured. Typical reporter genes include β-galactosidase or a fluorescent protein. Historically, lineage tracing was used most in the field of developmental biology. In its earliest application, lineage tracing was accomplished not by the expression of a fluorescent reporter, but through direct observation of embryos by light microscopy. This allowed the generation of 'fate maps' for cells from the one-cell stage until generation of germ layers.[1]

Subsequent advances in the field included application of dyes to the surface of an embryo in order to track the movement of groups of cells.[2] However, direct observation is limited to the study of small, transparent embryos, and as a result lipid-soluble dyes were injected into the embryo and later sectioned.[3] Ultimately, the introduction of genetic markers superseded dye experiments because of their advantages: they do not leak into neighboring cells and they are typically inherited by progeny where they are expressed at levels equal to that of the parental cell. These genetic markers were originally introduced by viral infection, direct injection, or transfection. Over time, techniques for lineage tracing have evolved to the point that they now possess remarkable sensitivity and offer sufficient resolution to track single cells, in real time. The subject of this review is the most commonly used lineage-tracing approach—genetic recombination.

GENETIC LINEAGE TRACING

The concept of genetic lineage tracing involves the expression of a recombinase enzyme in a cell-specific manner in order to activate the expression of a reporter gene. Cre recombinase (causes recombination of the bacteriophage P1 genome) has become an indispensible tool in the mouse model system, and it has recently been adapted for use in zebrafish.[4,5] Flippase is an alternate recombinase that has been used mostly in *Drosophila* and will not be discussed here. Cre recognizes a 34-base-pair nucleotide sequence called *loxP* that is not present in the mouse genome. When two loxP sites are oriented in a head-to-tail manner, Cre recombinase will excise the intervening DNA sequence while rejoining the ends together. This is very useful in turning genes on for lineage tracing. Normally, a reporter gene is located downstream of a

ubiquitous promoter, such as the Rosa26 locus.[6] If a strong transcriptional stop sequence, such as a triple polyadenylation sequence, is placed between two *loxP* sequences and located upstream of this reporter gene, this reporter gene will always be turned OFF. However, expression of Cre recombinase will cause deletion of the polyadenylation sequences, turning the reporter ON in that cell and all of its descendants.

By choosing cell-specific promoters for expression of Cre and combining it with a reporter, one can mark any cell population when that promoter becomes activated. An ever-growing variety of Cre driver mouse lines are available, and they have been created by several approaches: for example, by random integration of short promoter sequences or by much larger bacterial artificial chromosome transgenes. Of the two, bacterial artificial chromosome transgenics are more likely to faithfully recapitulate the endogenous expression pattern of the gene of interest, as unknown regulatory sequences are more likely to be included on the bacterial artificial chromosome. Knock-in approaches are also useful, and these are also highly likely to recapitulate endogenous expression patterns, as all nearby DNA sequences are intact at the locus of interest.

In many cases, it is useful to be able to inducibly mark a cell, and this is accomplished using a Cre enzyme fused to a modified form of the estrogen receptor (CreERt2), which causes Cre to be sequestered in the cytoplasm until it binds to its ligand tamoxifen, triggering nuclear translocation.[7–9] Both a codon-optimized version (iCreERt2) and a codon-optimized version fused to two ERt2 domains (iERCreER) at the N and C termini are now available and offer improved expression and recombination frequency.[10,11] Alternatively, a constitutively active Cre may be regulated by a drug-controlled promoter such as tetracycline. Tetracycline-regulated systems in particular are quite robust in mouse but are generally used for protein overexpression experiments.[12] They are less commonly used for lineage analysis because they may require three separate alleles in this application rather than two, making for a cumbersome (and expensive) breeding strategy. However this approach has been used with success along mouse nephron epithelia, for example, in which a Pax8-rtTA allele, which drives expression of the reverse tetracycline transactivator (rtTA) is crossed to a mouse line with an allele consisting of a tetracycline-response element driving Cre expression as well as a reporter mouse.[13] In this case, administration of doxycycline binds to rtTA present in Pax8-positive renal epithelial cells only, and the doxycycline-rtTA complex drives Cre expression and subsequently activation of the reporter.

Among the important issues to consider before choosing an appropriate reporter is the anticipated method of detection. The reporter chosen must possess sufficient signal-to-noise ratio to enable unambiguous detection (Figure 1). The *Escherichia coli* lacZ gene, β-galactosidase, has been used extensively and produces an intense blue color when incubated with the substrate analog X-gal. Yet β-galactosidase does not provide cellular resolution in thin cells, because the reaction product does not entirely fill long processes but is enriched in the cell body.[14] In addition, there is a lack of specific antibodies that would allow routine indirect immunofluorescence detection in all systems. Therefore, LacZ is not the best reporter to detect interstitial cells. Fluorescent reporters are now the norm, and in this case detection by epifluorescence is far superior to antibody-enhanced methods, which are subject to nonspecific binding of primary and secondary antibodies. Therefore, an adequate expression level of the reporter is critical. Although the Rosa26 locus drives strong expression of reporters in developing kidney, the expression level in adult kidney is reduced and may be inadequate (Humphreys, unpublished observations). Thus, straight Rosa26 reporters may not be ideal for experiments in adults. In such cases, reporters that have an additional CAG promoter knocked into the locus provide much stronger expression levels.

Older reporters exist in which viral promoters such as CAG were generated as random integration transgenics, rather than knocked into the Rosa26 locus.[15] These may suffer from mosaic expression as the integration locus may be inactive in certain cells, and this is an important limitation. The mT/mG reporter addresses this limitation because cells that have not undergone recombination express membrane-targeted tdTomato (mT), but after recombination they express membrane-targeted enhanced green fluorescent protein (eGFP) (mG, Figure 1a). In this way, mT expression verifies that cells of interest are capable of expressing the reporter allele.[16]

Fluorophores with strong epifluorescence also improve signal-to-noise ratio. Although eGFP and its variants continue to be quite useful, a new generation of fluorophores offer much brighter fluorescence. The most highly fluorescent reporter currently available is a tandem dimmer of Tomato (tdTomato), which is itself a mutagenized version of DsRed. tdTomato is nearly three times brighter than eGFP, is nontoxic to cells, and is photostable.[17] A Rosa26 knock-in reporter mouse expressing tdTomato is available, in which a strong CAG promoter has been inserted to increase expression levels[18] and a woodchuck hepatitis virus post-transcriptional regulatory element is present in order to further enhance mRNA stability (Figure 2).[19] These mice can be maintained as homozygotes. The primary weakness of this line is that tdTomato fluorescence is so strong that it can easily bleed into the eGFP channel.

CRITICAL STEPS IN LINEAGE TRACING

Nearly any kidney cell type can now be targeted through widely available Cre drivers and reporter strains, but it remains critical to characterize each system in order to interpret results accurately. For example, mosaic expression of Cre in target tissues can lead to incomplete recombination and false-negative results. This may either be due to silencing of the transgene in a subset of cells or due to low absolute expression levels. An even more serious problem is unfaithful expression outside of the cell type that should express Cre, causing false-positive results. Many Cre drivers are created with short transgenic promoters, and these may have