Reconstructing the lineage of cells is central to understanding development and is now also an important issue in stem cell research. Technological advances in genetically engineered permanent cell labeling, together with a multiplicity of fluorescent markers and sophisticated imaging, open new possibilities for prospective and retrospective clonal analysis.

**History and Concepts**

Cell lineage analysis is intimately connected with the emergence of developmental biology as a field of scientific research (see Galperin, 1998). In the mid-19th century, identification of cells as the basic units of life (Schleiden, 1838; Schwann, 1839) led to the realization that cells come from pre-existing cells (Virchow, 1858). Before the end of the 19th century, the work of Charles Whitman (1887) and later of Edmund Wilson (1892) on leech and annelid embryos led to the formulation of the term “cell lineage.” This early work inspired the Wood’s Hole School at the Marine Biology Laboratory in Massachusetts, where pioneering research in invertebrate embryos led to important concepts for lineage analysis. Thus E. Wilson viewed lineage in terms of the fate of cells and E.G. Conklin (1905), another major figure, made the distinction between determinate and indeterminate types of cleavage in ascidians, leading to the concept of invariant and noninvariant cell lineages. Breakthroughs in vertebrate fate mapping came from the systematic use of vital staining of groups of cells (Vogt, 1929) and from grafting experiments (Spemann and Mangold, 1924) in the amphibian embryo. In addition to embryological approaches, the work of A.H. Sturtevant, based on genetic studies initiated by T.H. Morgan and others on spontaneously generated mosaicism in insects, led to retrospective analyses in which cell lineage and gene function were associated (Sturtevant, 1929).

Many of the conceptual issues of today were evident when cell lineages were first explored. Lineage studies, then as now, aim to establish which cells, and how many cells, in the early embryo will give rise to a structure and, as development proceeds, from which part of a structure a substructure derives. These interrogations now extend to the origin of stem cells that permit the regeneration of an adult structure as well as its initial formation. Clonal analyses, which describe the derivatives of a single cell, provide insight into the mode of growth of a tissue and its regionalization with potential clonal boundaries (Lescroart et al., 2010) between compartments, or with segregation between distinct cell lineages, which do not necessarily correspond to distinct differentiated cell types but rather to topographical subdivisions. Analyses of clones can also provide information about cell death and proliferation, cell competition, cell movement and dispersion, and tissue polarity. Experimentation in a growing number of tissues and model organisms reveals the diversity of cell behavior that underlies progression along a lineage tree and has led to the elaboration of conceptual frameworks for cell lineage analysis (e.g., Garcia-Bellido, 1985; Petit et al., 2005; Stent, 1985).

In the context of embryonic development, many invertebrates have invariant lineages, meaning that a blastomere not only has a predictable future but also has a reproducible position and a defined group of neighbors from one individual to another. This is illustrated by C. elegans, for which a complete lineage tree has been defined (Sulston et al., 1983). In contrast, in the early mouse embryo, for example, more cell mixing takes place and cells in the inner cell mass of the blastocyst retain pluripotency and plasticity (Cockburn and Rossant, 2010). In the case of such noninvariant (regulative) development it is more challenging to analyze cell lineages.

Intertwined with the concept of lineage is that of cell commitment. Cell lineage follows the normal fate of a cell and its daughters, leading to the formulation of genealogical trees of cells with increasingly restricted cell fate choices as development proceeds. Unlike lineage, commitment can only be established by experimental challenge, such as ectopic grafting or in vitro manipulation, showing that the cell has acquired a restricted cell fate potential (e.g., Tam et al., 1997). As G. Stent (1985) pointed out, cell lineage plays a role in cell commitment by the unequal partitioning of cell determinants in daughter cells in successive cell divisions, as illustrated by the ascidian, Ciona (Nishida, 1987), and by the orderly placement of cells relative to intercellular signals as development proceeds, which is a major feature of vertebrate embryogenesis. It has become increasingly clear that even differentiated cells retain plasticity, as demonstrated by the spectacular phenomenon of induced pluripotency (Yamanaka, 2009). Caution should be exercised in equating cell fate restriction with gene expression. Characterizing lineage progression in these terms provides a genetic complement to cellular studies but can also lead to experimental pitfalls as discussed later.

Experimentalists today face many of the same dilemmas that confronted embryologists a hundred years ago—namely, how to label cells and subsequently analyze their contribution to the embryo based on the perdurance of the label, without perturbing the development of the organism. New technological developments now facilitate detailed analysis of complex situations.
In the following sections we shall discuss the current state of this art and future developments, where temporal as well as spatial regulation of the onset of labeling, simultaneous detection of several lineages, systematic labeling of all progenitors of a structure, visualization of the dynamics of lineage progression, and linking lineage to gene function are the underlying issues. In this era of genetic tools for cell tracing, we will focus on *Drosophila*, zebrafish, and mouse, with reference to avians, amphibians, and plants, as well as to the other invertebrate models that have provided important insights into lineage analysis. Approaches currently employed for following the history of a cell, which we discuss here, are summarized in Table 1.

**Prospective Lineage Analysis: Selection of Labeled Progenitors**

Prospective lineage analysis is a classic approach in which cell labeling is performed at a known position and stage and the contribution of the cell’s descendants to a structure is subsequently analyzed. In this context, cell fate mapping can be achieved by grafting experiments where labeled cells are introduced into the embryo and their subsequent contributions monitored. Following on from the pioneering work of Mangold and Spemann in amphibians (Spemann and Mangold, 1924) and of Waddington in the chick (Waddington, 1932), the analysis of chick-quail chimaeras, in which it is possible to distinguish nuclei between these closely related species, underlies important aspects of our understanding of cell fates in the amniote embryo, as exemplified by neural crest cell derivatives (Le Douarin and Barq, 1969). In an alternative, generally applicable approach, radioactive labeling of transplanted cells has been used (Weston, 1963) and was instrumental in mapping the heart-forming fields in the chick embryo (Rosenquist and De Haan, 1966). This approach was also used in challenging experiments on the mouse embryo to map the fate of cells in the epiblast (Beddington, 1981). Lypophlic carboxyamine dyes, such as Dil, which can be introduced into a small region of the embryo have been, and continue to be, used to trace groups of cells and examine cell movements. Less invasive than grafting experiments, these vital dyes intercalate into the cell membrane and are easily visualized. In early experiments with this method of cell marking, the migration pathways of neural crest cells and the temporal order in which they contribute to their derivatives were refined for the chick embryo (Serbedzija et al., 1989). Numerous fate maps have been established using this technique, including more recently in the lamprey (McCauley and Bronner-Fraser, 2003) and in cultured mouse embryos (Galli et al., 2008).

Single-cell labeling, which permits lineage analysis, is more challenging but can be achieved by microinjection. Classically, horseradish peroxidase (HRP) or dextran linked fluorescent dyes, which are too large to diffuse between cells, have been used as intracellular markers. Pioneering experiments using HRP pressure mediated microinjection in the leech embryo, where there is little cell migration (Weisblat et al., 1978), showed that teloblasts, which are the founder cells of segments, give rise to topographically invariant lineages that consist of different cell types. Interestingly, in this case, unlike that of *Drosophila* (Garcia-Bellido, 1985), morphological segment boundaries do not necessarily correspond to borders of clonal restriction. In the frog embryo, the progeny of different blastomeres were shown to contribute to distinct clonal domains with well-defined boundaries in the central nervous system (Hirose and Jacobson, 1979), although the descendants of a blastomere are not restricted to a single neural, or other, cell fate, indicative of global cell mixing (Moody, 1987). As distinct from amphibians, early planes of cleavage are not related to the plane of bilateral symmetry of the zebrafish embryo, and descendants of a single blastomere tend to remain associated initially (Kimmel and Law, 1985a) until they disperse at the onset of epiboly (Kimmel and Law, 1985b). During gastrulation, cell mixing decreases and tissue-specific lineages have been observed from this stage (Kimmel and Warga, 1986). In the mouse embryo, where there is extensive cell mixing, single-cell labeling by iontophoresis microinjection of HRP (pioneered by Balakier and Pedersen, 1982) in cells of the epiblast has led to fate maps in which the probability of descendants of a cell contributing to a particular tissue was determined (Lawson et al., 1991). In the absence of stereotyped lineages and despite geometrical differences, topological fate relationships at the stage of gastrulation are conserved between mammals, birds, amphibians, and zebrafish.

The advent of fluorescent proteins as markers (Shimomura et al., 1962) has had a major impact on fate mapping and cell tracking, as they are genetically encoded. Furthermore, fusion proteins, in which the fluorescent protein is targeted to the nucleus (e.g., H2B-GFP) or to the plasma membrane, provide clearer cellular resolution and additional information, such as mitotic status or cell shape dynamics. Thus, for example, microinjection of DNA encoding a fluorescent protein, GFP, demonstrated that a single cell in the chick somite is bipotent and revealed how its descendants progressively acquire a dermal or muscle cell fate (Ben-Yair and Kalcheim, 2005). In another example, microinjection of mRNAs encoding membrane-bound fluorescent proteins into a single cell of the inner cell mass of the mouse blastocyst, followed by time-lapse imaging in relation to a chromosomal marker, has shown how segregation between epiblast and primitive endoderm lineages is accompanied by extensive cell movement and, coupled with early markers of these cell types, supports the conclusion that primitive endoderm formation involves cell sorting and position-dependant induction (Meilhac et al., 2009). In this case, the characteristics of clones were used to test computer models of mechanisms for lineage segregation. In invertebrate models, too, fluorescent proteins are being used to track cells. In the leech embryo, an analysis based on injection of a plasmid encoding H2B-GFP now indicates a transition from tightly regulated to more stochastic cell division, pointing to a less black-and-white distinction between invariant and non-invariant lineages (Gline et al., 2009), as observed to some extent even in *C. elegans* (Schnabel et al., 1997).

Advances in understanding chromophore photochemistry have made it possible to engineer photomodulatable fluorescent proteins (see Platkevich et al., 2010), which have tended to replace caged molecules that require chemical synthesis, for marker activation. The value of such a caged dye was first shown in an experiment in *Drosophila* that revealed clonal restriction anteriorly but not posteriorly when the dye was activated at the site of establishment of an Engrailed 1-positive parasegment (Vincent and O’Farrell, 1992). In the last few years a range of photoconvertible fluorescent proteins, which undergo a spectral...
<table>
<thead>
<tr>
<th>Current Methods</th>
<th>Requirements</th>
<th>Merits</th>
<th>Limitations</th>
<th>Organism</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prospective</strong></td>
<td></td>
<td></td>
<td></td>
<td>M C F Z D I P</td>
</tr>
<tr>
<td>Dil</td>
<td>vital staining, embryo or explant culture</td>
<td>targeted, easier than microinjection</td>
<td>nonclonal, dilution, accessibility to cells</td>
<td>x x x x</td>
</tr>
<tr>
<td>HRP, dextran microinjection</td>
<td>micromanipulation, embryo or explant culture</td>
<td>clonal, targeted</td>
<td>dilution, invasive, accessibility to cells</td>
<td>x x x x x</td>
</tr>
<tr>
<td>DNA, RNA microinjection</td>
<td>micromanipulation, embryo or explant culture</td>
<td>clonal, targeted, amplification of the marker</td>
<td>dilution, invasive, accessibility to cells</td>
<td>x x x x x</td>
</tr>
<tr>
<td>Uncaging or photomodulation of a fluorescent protein</td>
<td>laser fluorescent microscopy, embryo or explant culture, injection of the marker or genetically modified line</td>
<td>targeted, no micromanipulation</td>
<td>dilution, phototoxicity, accessibility to cells</td>
<td>x x x</td>
</tr>
<tr>
<td>Genetic tracing / tissue-specific recombinase</td>
<td>genetically modified lines</td>
<td>noninvasive, permanent</td>
<td>nonclonal, dependent on gene/promoter expression and potential integration site effects</td>
<td>x x x</td>
</tr>
<tr>
<td>Transplantations</td>
<td>micromanipulation, labeled donor</td>
<td>permanent if genetic marker, easier than microinjection</td>
<td>nonclonal, invasive, accessibility to cells</td>
<td>x x x x</td>
</tr>
<tr>
<td><strong>Mosaic</strong></td>
<td></td>
<td></td>
<td></td>
<td>M C F Z D I P</td>
</tr>
<tr>
<td>Early chimaeras with cell mixing</td>
<td>micromanipulation, lines with distinct phenotypes</td>
<td>permanent, sparse labeling</td>
<td>nonclonal, invasive</td>
<td>x</td>
</tr>
<tr>
<td>DNA electroporation</td>
<td>electroporation, cell tracking</td>
<td>multicolor, sparse labeling</td>
<td>nonclonal, dilution, accessibility</td>
<td>x x</td>
</tr>
<tr>
<td>X-inactivation</td>
<td>genetically modified lines</td>
<td>spontaneous, permanent, sparse labeling</td>
<td>nonclonal</td>
<td>x</td>
</tr>
<tr>
<td>Multicolor genetic mosaics: MADM, twin-spot, brainbow/confetti</td>
<td>genetically modified lines, resolving color hues</td>
<td>multicolor, permanent, sparse labeling</td>
<td>nonclonal</td>
<td>x x</td>
</tr>
<tr>
<td><strong>Retrospective</strong></td>
<td></td>
<td></td>
<td></td>
<td>M C F Z D I P</td>
</tr>
<tr>
<td>Retrovirus</td>
<td>library of tagged retroviruses, isolation of cells for PCR/sequencing analysis</td>
<td>time-control, clonal</td>
<td>differential infectivity of cells, potential integration site effects</td>
<td>x x</td>
</tr>
<tr>
<td>Inducible recombinase: temperature, hormone or antibiotics</td>
<td>genetically modified lines, temperature shift or inducer molecule concentration, control background levels</td>
<td>time-control, dose control of clone frequency</td>
<td>toxicity, partial activity of the Cre, reproducibility</td>
<td>x x x x</td>
</tr>
<tr>
<td>Inducible transposon mobility: temperature, inducer molecule</td>
<td>genetically modified lines, temperature shift or inducer molecule concentration, control background levels</td>
<td>time-control, dose control of clone frequency</td>
<td>instability, potential integration site effects</td>
<td>x</td>
</tr>
<tr>
<td><strong>Spatially Random Labeling</strong></td>
<td></td>
<td></td>
<td></td>
<td>M C F Z D I P</td>
</tr>
<tr>
<td>Microsatellites</td>
<td>isolation of cells, sequencing analysis</td>
<td>systematic, spontaneous</td>
<td>large number of observations for statistical analysis</td>
<td>x</td>
</tr>
<tr>
<td>Mitotic recombination with a lacZ-like reporter</td>
<td>genetically modified line</td>
<td>systematic, spontaneous</td>
<td>large number of observations for statistical analysis</td>
<td>x x</td>
</tr>
</tbody>
</table>
change after exposure to activating light, has become available. These proteins can be photoconverted by confocal laser microscopy or even using regular fluorescence microscopes (Baker et al., 2010; Stark and Kulesa, 2007). In some instances, photoconversion by two-photon microscopy may be applicable (Hatta et al., 2006) for single-cell labeling, with the advantage of deeper penetration into the tissue and less phototoxicity. Issues in using fluorescent proteins for cell tracking include the rapidity and stability of photoconversion, the brightness of the fluorescence, and toxicity; it is important to test these parameters for the organism and developmental stage under study (e.g., Nowotschin and Hadjantonakis, 2009). The zebrafish embryo because of its accessibility, transparency, and rapid development particularly lends itself to fluorescent cell tracking. Temporal conversion of fluorescent proteins can be used for in vivo birthdating of tissue types. This is illustrated by the BAPTI system where a photoconvertible Kaede reporter is under the control of a neural promoter; after exposure to activating light early-born neurons are labeled red, whereas later-born neurons, not exposed to photoconversion, remain green (Figure 1). The BAPTISM system extends this to include an additional population-specific reporter. This analysis led to the conclusion that the specification and function of different classes of trigeminal sensory ganglia depend on the timing of neurogenesis (Caron et al., 2008).

With the approaches of prospective clonal analysis, fate maps can be drawn and partial lineages reconstructed. However, the use of microinjected markers or sequences encoding photomodulatable fluorescent proteins is limited by the problem of marker dilution at each cell division and thus is only applicable for short-term labeling experiments. Furthermore, their introduction is invasive, is often challenging technically, and may cause damage. A major interest of fluorescent proteins is that they can be employed after stable integration of their coding sequence into the genome by transgenesis or gene targeting and can therefore provide permanent cell labeling. This is classical in mouse and fly and is now becoming practicable in

<table>
<thead>
<tr>
<th>Current Methods</th>
<th>Requirements</th>
<th>Merits</th>
<th>Limitations</th>
<th>Organism</th>
</tr>
</thead>
<tbody>
<tr>
<td>4D Imaging</td>
<td>embryo or explant culture, high-resolution microscopy, computing capacity</td>
<td>dynamic, direct, comprehensive</td>
<td>limited developmental window, penetration, complexity of image analysis</td>
<td>M C F Z D I P</td>
</tr>
</tbody>
</table>

Table 1. Continued

This table is mainly based on papers discussed in the text. M, mouse; C, chick; F, frog; Z, zebrafish; D, Drosophila; I, other invertebrates; P, plants.

Figure 1. Live Imaging of Kaede Photoconversion in Zebrafish: An Example of Prospective Lineage Analysis
(A) Photoconversion of the Kaede fluorescent protein by exposure to UV light induces rupture of a covalent bond (adapted with permission from http://zeiss-campus.magnet.fsu.edu/articles/probes/highlighterfps.html).
(B) Early trigeminal sensory neurons born before the photoconversion appear yellow, due to the presence of the photoconverted red protein and the newly synthesized protein (green), whereas the neurons born later remain green.
(C) Schematic representation of the mode of action of this BAPTI system on cell lineage (Caron et al., 2008, adapted with permission).
zebrafish, based on transposon-mediated integration. Avian transgenic lines, based on lentiviral-mediated transgene integration, should also provide new tools for lineage analysis (Sato et al., 2010). Further improvements for directing transgene insertion in a range of species can be envisaged with zinc finger nucleases, meganucleases, or TALE nucleases (Christian et al., 2010).

Genetic Tracing

In order to follow the descendants of a cell, the recombinase approach to permanent genetic labeling by specific activation of a conditional reporter is widely used in mouse and fly and is now available in fish and has also been used for genetic cell tracing in Xenopus (Satoh et al., 2005). Recombinase activity should be rapid, efficient, and specific, although there can be problems with certain loci and with Cre toxicity, even in mice (Naiche and Papaioannou, 2007). This is a problem in Drosophila, where FLP recombinases are the preferred tools. Improved variants of FLP and Cre, together with the identification of specific target site variants (FRT, lox), have increased the efficiency and scope of these tools (Turan et al., 2011). In addition to reporter lines where a single marker is activated on recombination, switchable lines in which recombination removes or inverses a first fluorescent reporter cassette, so that a second cassette is expressed, permit marking of cell types before and after recombination (Muzumdar et al., 2007 in mouse; Boniface et al., 2009 in fish). In order to follow all cell derivatives a ubiquitously expressed regulatory sequence controlling the conditional reporter is required. In mouse, targeting to the Rosa26 locus is frequently used, with an additional CAG promoter sequence to cell population. Mouse Cre lines are extensively used to follow the descendants of cells that had expressed the Cre recombinase. A repertoire of Cre lines, which include flexible locus targeting, continues to be developed by consortia such as EUCOMM. This approach, described as “genetic inducible fate mapping,” was first employed in experiments where an Engrailed-Cre line was crossed with a β-actin-loxSTOPlox-lacZ line to fate map cells originating at the mouse midbrain-hindbrain constriction (Zinyk et al., 1998). In the fly model, the G-TRACE procedure (Evans et al., 2009) is based on the GAL4-UAS binary expression system (Brand and Perrimon, 1993), in which a sequence, encoding the FLP recombinase, is under the control of UAS regulatory elements that are targeted by the transcriptional activator GAL4, produced from another transgene with tissue-specific regulatory elements (Figures 2A and 2B). A strength of the Drosophila community has been the large collection of UAS/Gal4 lines and further resources, which integrate recombinase technology, is now becoming available. The GAL4-UAS system is now also being optimized for use in zebrafish, where lines are beginning to be established. As an alternative to the use of a recombinase, the Kaloop approach (Figure 2C) provides permanent cell labeling by autoinduction of Gal4 under the control of UAS in the reporter cassette independently of the tissue-specific promoter (Distel et al., 2009).

In all these approaches, genetic tracing of progenitors labels all the cells that had expressed the tissue-specific promoter driving Cre, precluding any distinction between different progenitors, and should be interpreted as the identification of structures that arise from a gene expression domain. Such genetic fate mapping in mouse, frequently confused with lineage analysis,
Developmental Cell

Perspective

has given rise to controversial results and needs careful control; transient gene expression in another progenitor cell population or unexpected later expression in cells of the tissue under study can confuse the analysis. This is exemplified by controversy over an experiment on genetic tracing of Tbx18-positive epicardial cells covering the mouse heart. In Tbx18-Cre;R26R embryos, myocardial cells expressing the lacZ reporter were observed, leading to the conclusion that the epicardium can give rise to myocardium. This was challenged by another group that had observed Tbx18 expression in some myocardial cells, which would trigger transcription of the reporter independently of expression in the epicardium (Christoffels et al., 2009). Temporal activation of Cre/Flo recombinases reduces the problem posed by misleading domains of expression.

Temporal regulation of marker gene expression is now an important facet of genetic tracing. In Drosophila, this can be achieved by activation, due to a change in temperature, of a heat-shock promoter (hsp) regulating FLP (Harrison and Perrimon, 1993), or by the use of temperature-sensitive versions of the GAL80 repressor (McGuire et al., 2003). In zebrafish, localized, or indeed cell-specific, activation of a heat-shock promoter has been successfully achieved by a laser-pointer-driven microheater (Placinta et al., 2009). The MAZe system in zebrafish (Figure 2D) depends on a self-excising hsp-Cre cassette that then brings a GAL4-VP16 cassette under the control of a ubiquitous promoter and leads to the activation of a UAS-driven fluorescent reporter in the same transgene (Collins et al., 2010), with resultant cell labeling. In mouse, regulation by temperature changes is not possible, and inducible systems involving antibiotics, such as the Tet system (Gossen and Bujard, 1992), or hormones, have been developed. The estrogen receptor (ER), with tamoxifen as a ligand, is widely used, mainly in the Cre-ER T2 version (Feil et al., 1997 in mouse; Mosimann et al., 2011 in fish). Tamoxifen administration by injection into the mother usually results in recombination within 6–24 hr in the mouse embryo (Hayashi and McMahon, 2002). In the case of the zebrafish embryo, immersion in a tamoxifen-containing medium, which can be washed out, permits rapid removal of the ligand. Imprecision in the timing of tamoxifen activation is a general problem; an ingenious method has been described recently for photostimulation of a caged form of tamoxifen in zebrafish, with the added advantage that a single cell can be targeted, with potential applications for clonal analysis (Sinha et al., 2010). As discussed later, dose control to give a low frequency of random recombination can be used to achieve clonal levels of cell labeling, permitting lineage analysis.

Genetic tracing reflects the activity of a promoter. However, mosaics, which are also based on genetic tools, provide wider possibilities for bona fide cell labeling.

Mosaics: Simultaneous Labeling of Several Progenitors

Simultaneous labeling of several progenitor cells may be useful to assess the variability of cell fate potential or to recognize differing cell fate choices after asymmetric division, or to monitor cell dynamics. Mosaics correspond to the association of genetically distinct cells, which reflect disparate, nonclonal labeling of progenitor cells. Mosaic embryos have provided important information about the origin of different cell types. Classic experiments using allophenic (chimaeric) mice, derived from morulae aggregation from different mouse lines, provided new insight into the polyclonal origin of a tissue or structure from a small number of founder cells, as exemplified by the melanocyte lineage that determines the ordered patterning of stripes of coat color (Mintz, 1965). Pioneering work based on the injection of ES cells into the blastocyst to create chimaeras (Gardner and Rossant, 1979) underlies more recent experiments. Thus, a mixture of ES cells, each with the Rosa26 locus targeted to express a distinct fluorescent protein, was introduced into wild-type blastocysts. The resultant mosaic embryos were analyzed to determine whether hematopoietic and endothelial cells in the yolk sac blood islands arise from a common hemangioblast progenitor. The authors concluded that each island had multiple progenitors and that the contribution of a single hemangioblast to both endothelial and hematopoietic lineages was a rare event (Ueno and Weissman, 2006). Creation of mosaics followed by time-lapse imaging has permitted cell tracking and fate mapping, as illustrated by limited electroporation of plasmids encoding H2B-EGFP and cytoplasmic DsRed into some cells of the chick epiblast. Multiphoton time-lapse microscopy showed the mechanism of the “polonaise” movements of cells that precede gastrulation; in this case the mosaic of positive and negative cells facilitated cell tracking (Voiculescu et al., 2007).

Genetic manipulations to produce mosaics have been classically used in Drosophila for clonal analysis. From pioneering work by Sturtevant (1929) on gynandromorphs resulting from spontaneous X-inactivation to current techniques for spatially and temporally controlled activation of cell markers, the coherent growth of cells in the Drosophila embryo has facilitated clonal analysis. Isolated clusters of labeled cells are generally assumed to be clonal.

More sophisticated reporters to mosaic labeling continue to be developed. For example, in mosaic analysis with double markers (MADM) (Zong et al., 2005), Cre-mediated recombination between homologous mouse chromosomes results in the generation of a complete coding sequence for GFP or RFP from chimaeric sequences containing partial sequences of the fluorescent proteins (Figures 3A and 3B). The use of split sequences also underlies the twin-spot system in Drosophila (Griffin et al., 2009). After mitosis, recombined chromatids may segregate to mark daughter cells differently, such that cells expressing either reporter come from a common progenitor.

With the development of spectral microscopy and of fluorescent proteins with a wide range of emission spectra, the number of markers that can be imaged simultaneously is increasing (Nowotschin et al., 2009). The Brainbow system, developed for mouse, depends on a stochastic choice between distinct recombinase target sites flanking a range of fluorescent markers in a transgene integrated in multiple copies. This leads to the generation of a spectacular mosaic of differently colored cells (Livet et al., 2007). With fluorescent proteins targeted to subcellular compartments, as well as with recombinase-mediated inversion of reporter sequences (Brainbow 2), the possible combinations that can distinguish cells are huge (<90). Although in the original paper the expression of the construct was restricted to neuronal cell types, a universal rainbow line, R26R-confetti (Snippert et al., 2010), coupled to existing specific Cre lines, increases the range of applications. This was used with a Cre-ER T2 under the control of Lgr5, which is expressed in stem
cells of the crypt of the mouse intestine. Tamoxifen induction of Cre at different time points, followed by mathematical analysis of cell patterns marked with the four randomly generated reporters, led to conclusions about stem cell turnover without asymmetric cell divisions. Stochastic adoption of stem or transit amplifying cell fates depends on neutral competition between cells. Such sophisticated mathematical analysis of clone distributions has also been applied to other tissues, and a general theoretical framework, which discriminates between patterns of long-term clonal evolution for distinguishing three classes of stem cell behavior, has been proposed (Klein and Simons, 2011). Adaptations of this system have now been described for Drosophila—d-Brainbow (Hampel et al., 2011) or Flybow (Hadjieconomou et al., 2011)—in which cell labeling as a result of different stochastic recombination events is linked to the UAS/GAL4 system to drive expression of the transgene reporter construct. In the d-Brainbow application, epitope tagged, as well as fluorescent proteins, were used, thus permitting both imaging and histological examination of fixed tissue. The Flybow system avoids the problem of Cre toxicity by employing Flp-mediated inversion of reporter sequences and has been used to address questions about the formation of neural network architecture (Figures 3C–3E). As in all mosaic analyses, the challenge is to sort out cells following the color code, which can be limited by the spectral separation of different combinations of fluorescent proteins and by the light microscopy resolution of subcellular localization.

Mosaics have opened the way to spectacular multicolor labeling of cells and have given insight into the polyclonal origin of tissues, their architecture, and the cell dynamics underlying tissue growth. The approaches discussed above, which integrate tissue-specific or temporal control, can also potentially be extended to introduce recombinases under cell-cycle control. Since they are generated genetically, an important application of mosaic approaches is that they can be combined with functional analyses, based on the use of mutated alleles.

Clonal Analysis and Gene Function

In order to relate clonal growth to gene function, mutant clones can be generated in a wild-type background or clonal analysis can be performed in a mutant background. Following the work of C. Stern (1936) on mitotic recombination, Flp-dependent interchromosomal recombination had been used in Drosophila to generate mutant clones that no longer express a marker (e.g., Xu and Rubin, 1993). An adaptation to produce positively marked clones, mosaic analysis with repressible cell marker (MARCM) (Figure 4A), leads to segregation of the mutant allele.
from the repressor GAL80, so that a UAS-driven fluorescent reporter is now activated in a mutated daughter cell and its descendants (Lee and Luo, 1999). Twin-spot MARCM combines the two approaches to follow sister cells (Yu et al., 2009). In this case, repressors are microRNAs that target UAS-dependent markers and are lost after mitotic recombination (Figure 4B). This reduces the delay of MARCM derepression, which otherwise depends on GAL80 decay. A potential difficulty is to determine when the mutation becomes effective, depending on the perdurance of the endogenous protein. Ideally, one would like to study lineage progression within clones of mutant cells. This can be achieved in Drosophila by combining the MARCM system with an independent Flp-induced recombination event that activates expression of a lacZ reporter (Figure 4C), so that two

Figure 4. Schematic Representations of Different Approaches to Tracing Cells in Mutant Clones in Drosophila

(A1) The MARCM system depends on the elimination, by mitotic recombination, of the Gal80 transgene, which lies on the same chromosome arm as the mutant allele (X). This results in their segregation and thus the activation of the GFP reporter in the homozygote mutant cells (Lee and Luo, 1999).

(A2) Transient heat-shock activation of the Flp recombinase results in reporter expression in mutant cells (Lee and Luo, 1999).

(B1 and B2) In twin-spot MARCM, the two systems are combined (see also Figure 3A), in this case using microRNAs (miR) as repressors of GFP or RFP expression (Yu et al., 2009).

(C1 and C2) To trace lineage progression in mutant clones, the MARCM system of GFP activation is combined with the system of heritable expression of lacZ (Harrison and Perrimon, 1993). A second reporter (lacZ) is activated after a second heat-shock-induced recombination that brings it under the control of a ubiquitous promoter (Perdigoto et al., 2011).
sequential recombination events can generate clones of β-galactosidase-positive cells within GFP marked clones of mutant cells (Perdigoto et al., 2011). In this way, Notch signaling in the lineage progression of intestinal stem cells has been shown to restrict their self-renewal as well as affecting the later stage of terminal differentiation.

Mosaics of cells with distinct genotypes also provide insight into cell-cell interactions and the importance of cell competition in selecting the progenitor of tissues and organs (Morata and Ripoll, 1975). For tracing the descendants of mutant cells in heterozygote mice, the MADM system can be used with distinct fluorescent reporters for wild-type and mutant alleles. For this purpose, a new reporter line, MADM-11, has been developed to generate clones mutant for a gene located on chromosome 11. With a Cre recombinase under the control of Emx1, which is expressed in cortical progenitors of the forebrain, it was shown that components of the Lis1/Nde1/14.3.3ε complex, which is defective in lissencephaly syndromes, have distinct cell-autonomous functions during different stages of neuronal migration (Hippenmeyer et al., 2010). The approach in which differently marked ES cells are introduced into a mouse blastocyst to create mosaics can be combined with functional studies on mutant ES cells to investigate the cell-autonomous roles of a gene. Alternatively, with a specific Cre, genetic tracing of clones can be achieved. Using a Fk1-Cre integrated into the genome of the reporter ES cells it was shown that all the endothelial cell derivatives in the blood islands were derived from progenitors expressing robust levels of Fk1, whereas most hematopoietic cells were not (Ueno and Weissman, 2006). Mosaics are also employed to perform clonal analysis on a mouse mutant background. Experiments in which insertion of a GFP reporter into the X chromosome resulted in random X-inactivation made it possible to follow cell dynamics in the developing limb bud by live imaging in a Wnt5a mutant embryo (Gros et al., 2010). Classic single-cell microinjection into the mouse epiblast has also been carried out with mutant embryos, for example to show that Otx2 is not required for proliferation of the visceral endoderm lineage but is essential for anteriorly directed cell movement (Perea-Gomez et al., 2001).

In addition to examining the effects of mutations on cell fate choice and associated cell behavior, a related challenge is to integrate the cellular data with dynamic gene expression, to understand how genes are expressed or repressed during progression along a lineage tree. The use of reporters of gene expression with limited stability, such as the fusion protein H2B-GFP (Plasa et al., 2008) or destabilized GFP (Harper et al., 2010), permits reliable imaging of gene expression dynamics. Integrating quantitative data on transcription factor kinetics with subsequent lineage patterning is now realizable. Thus monitoring the nucleocytoplasmic movement of Oct4 fused to a photoactivatable GFP has demonstrated that differences in Oct4 kinetics predict the future identity of mouse blastocyst lineages (Plachta et al., 2011).

Modeling the emergence of different cell types in a lineage is an emerging theme, as, for example, in C. elegans, for which a complete lineage tree has been reconstructed and regulatory genes have been identified. Such a predictive model gives an indication of the number of regulatory factors required for recapitulating the lineage, the synergistic variation of factors, and where, in the cell lineage tree, asymmetry might be controlled by external influences (Larsson et al., 2011). The superposition of experimentally determined gene regulatory networks on cell lineage is beautifully illustrated by pioneering work on the sea urchin embryo, where lineage is mainly invariant and early lineage segregation has been examined on a cell-by-cell basis in terms of transcriptional regulation and cell signaling. In this way, for example, the endoderm gene regulatory network has been defined up to the midblastula stage, giving new insight also into the progressive segregation of endodermal from mesodermal lineages (Peter and Davidson, 2010).

In most of the mosaic analyses previously discussed, the lineage is preidentified by the use of tissue-specific regulatory sequences and in this respect is therefore prospective. The interpretation of mosaics in terms of clones is limited to local events when growth is coherent and clusters of cells can be considered as clonal units. In retrospective clonal analysis, it is possible to lower the frequency of cell labeling to reach clonal conclusions, even when growth is dispersive and to do this on the scale of the whole organism.

Retrospective Clonal Analysis: Systematic Analysis of All Progenitors of a Structure

Prospective lineage analysis depends on a preconceived idea about the progenitor cell population. Preidentification of the potential stage and location of the progenitors to label is required. However, this is not always known and potentially restricts conclusions on lineage by not considering other counterintuitive options. In contrast, retrospective approaches to cell lineage depend on analyzing labeled cells at the end point of the experiment and deducing their interrelationships and previous history. Retrospective clonal analysis based on the random genetic labeling of progenitors at a low frequency constitutes a progenitor screen and permits the systematic analysis of the potential of any progenitor to colonize a particular structure. We distinguish two kinds of retrospective clonal analyses, depending on whether labeling is random in space, but with temporal control, or random in both space and time.

Spatially Random Labeling

Clones that are induced at a particular time but result from random spatial labeling have been produced in Drosophila by X-irradiation induced recombination, pioneered by H.J. Becker (1957), and subsequently employed to generate, in a heterozygous Minute mutant background, Minute-positive cells with a cuticular marker, which tend to outgrow their mutant neighbors (Garcia-Bellido et al., 1973). The distribution of clusters of such cells demonstrated the existence of internal demarcation lines in the wing disc, which the clones did not cross. This classic work led to important concepts and definitions, including those of clonal compartments and clonal boundaries (Garcia-Bellido, 1985). In the plant kingdom, elegant experiments on genetic variegation have manipulated transposon-mediated gene silencing. For example, a change in temperature led to low-frequency mobility of a transposon in the promoter region of the Pal gene, which encodes a red pigment, resulting in restoration of gene function to give red sectors (clones) on the ivory background of an Antirrhinum petal. In such experiments, the sequence of lineage restrictions in the developing floral meristem has been revealed (Vincent et al., 1995).
quantitative analysis of sector parameters coupled to computer modeling led to conclusions on the growth parameters that are essential to give shape to the flower petal (Rolland-Lagan et al., 2005).

In amniotes, infection with replication-defective retroviruses that integrate into the genome provides a spatially random labeling of progenitors at a defined time. This approach, first described for the mouse (Sanes et al., 1986), was refined for clonal analysis by the development of libraries of individually marked retroviruses, where each member encodes a reporter and has a DNA tag (Golden et al., 1995). The complexity of the library permits evaluation of clonality between labeled cells, based on the presence of the same tag, identified by PCR. This approach has been extensively used in birds and mammals, especially for characterizing lineage segregation in the central nervous system, as in the early demonstration that clonal derivatives contribute to more than one major subdivision of the telencephalon (Walsh and Cepko, 1992). In a recent adaptation, used for clonal analysis of blood cell types, infection with a barcoded retroviral library carrying a fluorescent marker was followed by separation of individual circulating cells by flow cytometry and sensitive sequence-based characterization of clonally related cells (Gerrits et al., 2010). Potential problems for random retroviral labeling are integration site effects and variable infectivity, illustrated by murine retroviruses that only infect proliferative cells, a limitation partially overcome by lentiviruses based vectors. As an alternative to libraries, low-level infection of GFP encoding retro- and lentiviruses has been used to mark single cells—for example, recently in the zebrafish brain, to show that whereas neuroblasts undergo a limited amplification, single radial glial cells self-renew and generate different cell types, thus behaving as bona fide stem cells in vivo (Rothenaigner et al., 2011).

Inducible recombinase systems based on the use of a ubiquitous reporter also provide random spatial labeling of progenitors. Adjusting the duration and temperature difference of the heat shock or the dose of tamoxifen can result in control of the frequency of labeling, to permit clonal conclusions. This approach (Figures 5A–5D), in which low doses of tamoxifen were administered to a CMV-CreERT2, R26R mouse line, was instrumental in showing the organization of stem cells in the matrix of the hair follicle and the mode of growth of their derivatives (Legué and Nicolas, 2005). In this example, statistical analysis was necessary to assess the probability of independent labeling events and to conclude on the clonal relationship between two labeled groups of cells. In a refinement of this method, two inducible reporters, R26R (lacZ) and R26R-EYFP, were used together to help to distinguish clonal events (Arques et al., 2007). Sophisticated quantitative analysis of such clonal patterns, coupled to computer modeling, has shed new light on the mechanism of limb bud growth (Marcon et al., 2011). Retrospective examination of lineage in this way, by activation of reporter expression at different time points, extends the potential for precise temporal reconstruction of lineage trees.

**Spatially and Temporally Random Labeling**

Other approaches to retrospective clonal analysis are random in both space and time and therefore encompass the complete history of a lineage. The accumulation of random somatic cell mutations during normal development provides an endogenous marker of cell lineage. Thus, analysis of mutations in microsatellite DNAs, at the single-cell level, using the new sequencing technologies, has led to the construction of mammalian lineage trees for a number of tissue types with easily isolated cells, such as the blood (Wasserstrom et al., 2008). This method is labor intensive and requires sophisticated computational analyses but is noninvasive, with the advantage that it is also applicable to human material.

Another approach depends on the introduction, as a transgene or targeted to an endogenous locus, of a nlacZ reporter sequence, rendered nonfunctional by a duplication that introduces a STOP codon into the β-galactosidase coding sequence. A rare, random event of intragenic recombination will generate a functional nlacZ reporter, which is then transmitted genetically to the descendants of the cell. This results in clonally related labeled cells, which are detectable when the recombined nlacZ lineage tracer is expressed (Figures 5E and 5F). The choice of regulatory sequences controlling reporter expression determines the tissue analyzed at the end point but does not condition the genetic labeling of the progenitor cells that give rise to it. The rarity of the event makes clonal analysis possible (Bonnerot and Nicolas, 1993). Collections of embryos are generated, in which the frequency of labeling, in the structure under consideration, is determined. To establish clonality, statistical analyses are required, based on the frequency of observations, as, for example, the fluctuation test of Luria and Delbrück (1943), which estimates the probability of one or more than one recombination events. The different types of clones that result from random labeling can be divided into groups based on their characteristics such as size, spatial distribution, and cell type. When similar clones are observed it can be assumed that the library of clones has reached saturation. From the collection of clones, derived from progenitors that have undergone recombination at different stages, the temporal history can be reconstructed, based on the premise that subclones have more restricted cell fate potential than parental clones. In addition to reconstruction of the lineage, its number of founder cells, and diversification into sublineages, important aspects of cell behavior such as the formation of clonal boundaries, or asymmetric stem cell versus symmetric proliferative modes of cell division, can also be deduced from the properties of clones within the library (Nicolas et al., 1996; Petit et al., 2005). An example is provided by analysis of clones in embryos of an α-cardiac actinnlacZ mouse line, which led to the demonstration of two myocardial cell lineages, which segregate early, with distinct and overlapping contributions to different parts of the heart (Meilhac et al., 2004). Subsequent analysis, using the same mouse line, established sublineages, within the second myocardial lineage, that contribute to different parts of the arterial pole of the heart and also to different skeletal muscle groups in the head (Lescroart et al., 2010) (Figures 5G–5L). The nlacZ approach has resulted in new lineage insights for many tissues in the mouse, including segregation of the germ layers during gastrulation (Tzouanacou et al., 2009), and should be applicable to other species. A similar reporter, based on duplication in the β-glucuronidase gene, has been used for clonal analysis in plants (Swoboda et al., 1994).

The strength of random retrospective clonal analyses is to reconstruct lineage trees over an extended time scale and to understand the mode of regionalization of a structure, by target-
events that precede the observations can be controversial. To gain direct access to the dynamics of lineage progression, live analyses are required, associated with successive observations at shorter time intervals.

Four-Dimensional Imaging of Lineage Progression

A grail of lineage analysis is the complete four-dimensional (4D) imaging of cells in vivo. This was pioneered in the chick, in which time-lapse imaging was first attempted over 80 years ago.

**Figure 5. Retrospective Clonal Analysis in Mouse**

(A–D) Inducible clonal analysis based on recombination of a conditional ROSA26 reporter (A) activated by low doses of tamoxifen (B). Clones in different layers (color coded; IRS, inner root sheath) of the hair follicle originate from different domains (red arrowheads) of the matrix which is a source of stem cells (C and D). (Legue and Nicolas, 2005, adapted with permission).

(E–L) Random retrospective clonal analysis by the lacZ approach (E) The lacZ reporter is rendered nonfunctional by an internal duplication, which can spontaneously recombine into a functional lacZ gene, at a low frequency.

(F) Random generation of an lacZ-positive clone (cells outlined in blue), which is detectable in the expression domain of the α-cardiac actin promoter, i.e., in cardiac and skeletal muscles (full blue circles indicate β-galactosidase [β-gal]-positive cells).

(G and H) Examples of β-gal-positive clones with an exclusive contribution to region 1 (outflow tract) or 3 (left ventricle) of the myocardium, indicative of lineage segregation. E8.5, embryonic day 8.5 (followed by clone number).

(I) The contributions of the first and second myocardial lineages, based on the analysis of 3,629 embryonic heart tubes, are summarized in red and green (Meilhac et al., 2004).

(J and K) Examples of β-gal-positive clones of the second myocardial lineage colonizing both skeletal muscles of the head and myocardium of the heart, taken from a collection of 2,223 fetuses.

(L) The lineage contributing to head muscles derived from the first branchial arch (1BA), which also contributes to the right ventricle (RV), is shown in blue, while the lineage that contributes to second branchial arch-derived head muscles (2BA) and to the outflow tract (OFT) is represented in pink (Lescroart et al., 2010).
cell cycles and appears to account for clonal restriction in neural cell fate (Kimmel et al., 1994). Time-lapse imaging of cortical slices in the mouse embryo, after in utero infection with GFP-expressing retroviruses, showed that radial glia generate neurons by asymmetric cell division (Noctor et al., 2001). However, with the exception of experiments on invertebrates, only fragmentary information on specific lineages has been obtained, mainly limited by the cell-labeling procedure.

For more systematic lineage reconstructions, pioneering experiments in plants, where the absence of cell migration and also of apoptosis facilitate cell tracking, have used fluorescent markers expressed in all cells, with a subcellular resolution, targeting the membrane or chromatin or revealing cell-cycle stages. In this case, confocal microscopy with long-term (12 days), as well as short-term, imaging was employed to analyze the development of the flower primordium from the meristem in Arabidopsis. Image registration algorithms were developed to assist lineage reconstruction (Reddy et al., 2004). In the plant field, the challenge of quantitative analysis of growth parameters over time is being met by new tools for image processing and reconstruction to track cell lineages (Fernandez et al., 2010). During animal development, confocal and multi-photon microscopy are currently extensively used for imaging cells. However, problems of light scattering and resolution are particularly critical for cell tracking. With the current interest in stem cells in the adult, as well as during development, accessing cells that are located deep within an organism can be a major problem. Advances in light sheet microscopy (e.g., SPIM, DSLM), successfully used on zebrafish embryos at early (Keller et al., 2008) or late (Swoger et al., 2011) developmental stages, hold out new promise for minimally invasive, high-resolution images with good penetration depth and fast acquisition (Huisken and Stainier, 2009). By following fluorescently labeled nuclei, with an automated image segmentation procedure, Keller et al. (2008) provide a resource of “digital embryos,” for lineage analysis over 24 hr, from early cleavage stages until the onset of organogenesis. Another new development, which is based on label-free multiphoton technology with spiral scanning to optimize resolution, penetration, and photoperturbation, has led to complete lineage reconstruction of zebrafish early development, up until the 1,000-cell (blastula) stage (Olivier et al., 2010) (Figure 6). This remarkable technical feat exploited the intrinsic optical nonlinear properties of the sample (harmonic generation)—namely, the oriented microtubules of the spindle and aqueous/lipidic surfaces such as membranes—together with two-photon excitation of a fluorescent chromosomal marker encoded by a transgene. This analysis required sophisticated processes of image acquisition, as well as algorithms for automated lineage reconstruction, which nevertheless still relied on time-consuming visual verification. A challenge for these new approaches is the requirement for higher-resolution multicolor cell imaging and for optimization of algorithms for fast and automated image reconstruction in dense environments, so as to allow unambiguous cell tracking. A practical problem for extending long-term imaging to later stages is the need to immobilize the animal without impeding development. It is not clear how far 4D analysis will progress to map cell behavior from the outset of development, but it also provides the potential to extend fragmentary imaging of later cell lineage choices into more comprehensive documentary films of cell history.
In conclusion, the technological developments discussed here have opened up new horizons. Genetic tools, which are becoming available for an increasing number of organisms, have solved the old problem of marker dilution at cell division and have introduced sophisticated methods for spatio-temporal targeting of labeling. Developments in the range of fluorescent markers now permit direct and multicolor observations, which can be orchestrated at will by genetic engineering. The classical problem of clonality of the labeling is addressed either by combining several markers or by lowering the frequency of labeling. Advances in microscopy are crucial for recording clonal data with increasing resolution in four dimensions. Sophisticated computational methods are being developed to analyze the large data sets generated by clonal analyses and to provide in-depth understanding of the cellular mechanisms leading to the observed clonal patterns. A challenge for the next decade is to grasp the significance of changes in cell behavior followed at a single-cell level and to integrate the cellular with the molecular dimension to understand lineage choices and lineage progression. In the future, in vivo imaging and genetic manipulation of markers will be widely applicable to the diversity of species already apprehended in Evo/Devo type studies, no doubt leading to unexpected conceptual lineage developments and revealing the cellular aspects of evolutionary “tinkering” (Jacob, 1977).

ACKNOWLEDGMENTS

We thank P. Herbomel, E. Hirninger, J.-F. Nicolas, and F. Schweisguth for insightful discussions and anonymous reviewers for comments. Work on clonality analysis in the Buckingham lab is supported by the Institut Pasteur, insightul discussions and anonymous reviewers for comments. Work on lineage analysis in the Buckingham lab is supported by the Institut Pasteur, the CNRS (URA 2578) and by the EU through the CardioCell (FP7-HEALTH-2007-2.4.2-5) project. S.M.M. is an INSERM research scientist.

REFERENCES


Perspective
Legue studies of gene function in neuronal morphogenesis. Neuron
113
132
Larsson, J.K., Wadstro¨mer, N., Hermanson, O., Lendahl, U., and Forchheimer,
¨
mer, N., Hermanson, O., Lendahl, U., and Forchheimer,


