Published in final edited form as:

Evol Dev. 2010 January ; 12(1): 84–94. doi:10.1111/j.1525-142X.2009.00393.x.

Phylogenetic analysis of developmental and postnatal mouse cell lineages

Stephen J. Salipante^a, Arnold Kas^b, Eva McMonagle^b, and Marshall Horwitz^{a,b,*}

^aDepartment of Genome Sciences, University of Washington School of Medicine, Box 358056, Seattle, WA 98109, USA

^bDepartment of Pathology, University of Washington School of Medicine, Box 358056, Seattle, WA 98109, USA

SUMMARY

Fate maps depict how cells relate together through past lineage relationships, and are useful tools for studying developmental and somatic processes. However, with existing technologies, it has not been possible to generate detailed fate maps of complex organisms such as the mouse. We and others have therefore proposed a novel approach, “phylogenetic fate mapping,” where patterns of somatic mutation carried by the individual cells of an animal are used to retrospectively deduce lineage relationships through phylogenetic inference. Here, we have cataloged genomic polymorphisms at 324 mutation-prone polyguanine tracts for nearly 300 cells isolated from a single mouse, and have explored the cells’ lineage relationships both phylogenetically and through a network-based approach. We present a model of mouse embryogenesis, where an early period of substantial cell mixing is followed by more coherent growth of clones later. We find that cells from certain tissues have greater numbers of close relatives in other specific tissues than expected from chance, suggesting that those populations arise from a similar pool of ancestral lineages. Finally, we have investigated the dynamics of cell turnover (the frequency of cell loss and replacement) in postnatal tissues. This work offers a longitudinal study of developmental lineages, from conception to adulthood, and provides insight into basic questions of mouse embryology as well as the somatic processes that occur after birth.

INTRODUCTION

The comprehensive embryonic development of the small, transparent nematode *Caenorhabditis elegans* has been described in that organism's “fate map” (Sulston et al. 1983), a diagram of lineages that depicts the derivation of individual cells during development, and conversely, that allows embryonic origins of specific body structures to be retrospectively traced (Clarke and Tickle 1999; Stern and Fraser 2001). Construction of the fate map was made possible by *C. elegans*’ transparency and rapid embryogenesis, permitting microscopic observation of all cell divisions.

In contrast to the worm, far less is known about development in complex animals like mice and humans, which do not lend themselves to observational approaches. Fate maps for more complex organisms must be inferred indirectly: typically, individual cells are prospectively tagged with dyes (Honig and Hume 1989) or genetic reporters (Eloy-Trinquet et al. 2000; Zong et al. 2005; Livet et al. 2007) so that their descendants can be identified at later stages. Although

cell labeling has proven valuable, it does not provide nearly as much information as direct observation; it can demonstrate only that a population of cells shares a common ancestor, without resolving the hierarchy of lineages through which they are related (Salipante and Horwitz 2007).

In light of this limitation we (Salipante and Horwitz 2006) and others (Frumkin et al. 2005) have proposed a novel approach to tracing cellular lineages, “phylogenetic fate mapping,” that, in principle, could be used to derive a fate map of every cell in a mouse or a human. Mutations inevitably accumulate each time a cell divides (Drake et al. 1998; Jackson and Loeb 2001), acting as a passive record of lineages wherein cells with the most closely related patterns of somatic mutation share the most recent common ancestry. By cataloguing genomic variation within the individual cells of an organism it is therefore possible to retrospectively infer their lineage relationships using phylogenetics algorithms, in the same way that similarities and differences in the genomes of species have been used to deduce population structure in evolutionary biology (Felsenstein 1981, 1988).

It is not yet feasible to sequence the entire genome of a single cell, so our approach currently necessitates examining mutational hotspots that provide sufficient information for reconstructing lineages. In prior work we demonstrated that polyguanine repeats are well suited for use in phylogenetic fate mapping (Salipante and Horwitz 2006), because they are prone to mitotic mutations that alter their length. Examining those sites has allowed successful reconstruction of a phylogeny of cultured mouse fibroblasts (Salipante and Horwitz 2006) and “proof-of-principle” fate maps of cells from normal adult mice (Salipante and Horwitz 2006; Salipante et al. 2008). Others have performed related studies, albeit using more mitotically stable short tandem repeat sequences coupled with a mutator genetic background (Frumkin et al. 2005; Wasserstrom et al. 2008a).

Here, we have generated a phylogenetic fate map of nearly 300 cells from an individual mouse, making it possible to infer the longitudinal lineage history of its cells, from fertilization to adulthood. The fate map serves as an initial survey of mouse development, offering insight into fundamental questions of embryology. Secondly, because phylogenetic fate maps have the advantage of being retrospective, we have used the data to investigate how cells are lost and replaced in postnatal tissues.

MATERIALS AND METHODS

Cell isolation and culture

Individual cells were cultured from a 7-week-old male, homozygous Immortomouse (Charles River). The Immortomouse is a transgenic strain, which allows for conditional immortalization of cells through selective induction of a SV40 T-antigen oncogene. Briefly, single cells were liberated from tissues, and clones were grown to confluence on 10 cm culture dishes. Colonies used for internal controls were each split to two separate culture dishes at approximately 64-cell stage and grown separately (“a” and “b”). DNA was purified using QIAmp DNA Mini kit (Qiagen). A full description of cell isolation, identification, and culture techniques are available in supplementary methods.

Genotyping

Oligonucleotides (purchased from ABI for NED-labeled primers and Operon for all other labels) are listed (supplement, primer sequences). Reverse primers carried the “pigtail” sequence 5'-GTTTCTT-3', which prevents genotyping artifacts (Brownstein et al. 1996). Five microliters of PCR amplifications containing roughly 6 ng genomic DNA each were carried out for 42 cycles using Taq DNA polymerase (Qiagen), and PCR fragments were resolved with

an ABI PRISM 3730xl Genetic Analyzer. Three technical replicates were performed for each sample, and a high degree of concordance was observed between independent replicates. Electropherograms were manually analyzed with ABI Gene-Mapper 4.0 software using the PeakSeeker approach (Thompson and Salipante 2009).

Some genotypes were ambiguous, appearing “half-way” between two interpretations, and it was not possible to confidently assign the length of both alleles. We attribute these ambiguous genotypes to genomic variation within the population of harvested cells, because genetic drift and/or competition between cells carrying new mutations likely occur during in vitro expansion. We coded alleles that could not be confidently typed as “missing” data (X).

Phylogenetic reconstruction

Phylogenetic analyses were performed using the Bayesian method as implemented in MrBayes 3.1.1 software (Ronquist and Huelsenbeck 2003), as described previously (Salipante et al. 2008), except that consensus trees were calculated using 10,000 trees after convergence was reached.

To estimate when in cells’ lineage histories they last shared a common ancestor, we compared the estimated number of cell divisions occurring before the cells’ inferred common ancestor to the number occurring after that ancestor. We calculated the branch length from the estimated zygote to the most recent common ancestor, and expressed it as a fraction of the total distance from the zygote to the cells. Because the branch length for cells following the common ancestor may not be equal, we calculated branch length after that point as the average for each of the two cells.

Network analysis

The purpose of the network analysis was to determine whether cells from a given tissue have a significant number of closely related relatives in other tissues. A full description of the algorithm is available in supplementary methods.

Mitotic history of individual tissues

We computed the number of mutations distinguishing each cell from the estimated zygote according to the pairwise distance matrix described previously.

Mutations are stochastic, thus, for a population where every cell has undergone the same number of divisions from the zygote, the number of mutations per cell is expected to follow a normal distribution having a standard deviation equal to $\sqrt{(\text{mean mutations})}$. The distribution of mutations carried by cells in each tissue was plotted, and how well the data fit that model was determined using the Anderson-Darling normality test.

We considered two alternative models for the distribution of mitoses. One was a normal distribution where the mean and the variance were separate parameters, and the other represented two subpopulations where the parameters corresponded to the number of cell divisions in each population and the fraction of cells contained within. For either model, we selected the parameters that best fit the real data using a χ^2 -test. Parameter selection was performed multiple times, using several random starting points, but each attempt gave nearly identical results. The corresponding “best fit” distribution of mutations under each model were compared with the actual data using a χ^2 -test in order to evaluate how well it fit.

RESULTS

Establishing cell lines

To construct our fate map, we analyzed single cells from various tissues of an individual mouse. So that sufficient DNA can be obtained from one cell for genotyping multiple markers by PCR, the genome of that cell must first be amplified. We previously showed this can be achieved by subjecting cells to in vitro “whole genome amplification” techniques (Klein et al. 1999; Salipante and Horwitz 2006), or by growing single cells in culture after their removal from a mouse (Salipante et al. 2008). Here, we have clonally expanded cells by culturing them, because mitosis replicates DNA more faithfully than in vitro techniques (Dietmaier et al. 1999; Hughes et al. 2005). Although individual cells are likely to accumulate additional mutations during the growth period, the majority of cells in the expanded population will not be mutated at any given locus, and DNA obtained from them represents an “average” genotype approximating the founder cell (Salipante and Horwitz 2006; Salipante et al. 2008). In effect, by passing tissues through a single-cell bottleneck, we preserve patterns of mitotic mutation existing at the time of isolation.

To permit sustained growth of clones, and to broaden the spectrum of cell types that could be grown in culture, we utilized a transgenic mouse expressing an inducible, temperature-sensitive SV40 T-antigen (Jat et al. 1991; Vicart et al. 1994). Under permissive conditions, the immortalizing T-antigen gene becomes active, facilitating the growth of primary cells. We isolated and cultured 291 clonal cell lines from various tissues (supplement, genotypes table), including vascular endothelium, preadipocytes, myoblasts, muscle satellite cells, kidney podocytes, bone marrow stromal cells and fibroblasts from lung, spleen, and muscle.

Genotyping

Following clonal expansion, we purified DNA from each of the cell lines and used PCR-amplification and capillary electrophoresis to determine the length of polyguanine tracts from 324 distinct autosomal markers (Fig. 1A). A length-altering somatic mutation in at least one clone was detected for 249 of the markers (76.9%). Most mutations were a single base insertion or deletion involving one allele, although larger length changes were observed. For 154 of the loci we found only one type of mutant genotype across all cells, while 95 loci exhibited two or more different mutant genotypes in various individual cells (Fig. 1B), implying that a subset of markers underwent multiple mutations during mouse development. Enough mutations were detected to uniquely distinguish all but a few isolates based on their pattern of somatic mutation (supplement, genotypes table), and only four clones did not have mutations which distinguished them from the zygote (Fig. 1C).

Establishing the genotype of the zygote

We estimated the genotype of the zygote empirically using large polyclonal mixtures of cells. Because somatic mutations are rare, unique polymorphisms in different lineages should not be distinguishable in such a sample, and its genotype thereby approximates that of the fertilized egg. We therefore purified DNA from two different large (approximately 0.5 g) tissue samples (arbitrarily selecting liver), and included them in our lineage reconstruction. Alternatively, we approximated the genotype of the zygote computationally by determining the most frequently occurring genotype at each locus. Reassuringly, all three estimations yielded similar results: the two experimental samples did not contain polymorphisms which distinguished them from one another, and the computed root differed from the experimental samples at only one allele of a single marker. For further analyses, one of the experimentally estimated samples was assigned as the root.

Phylogenetic analysis

A Bayesian phylogenetic reconstruction (Fig. 2A and phylogeny Newick) resulted in a tree with internal structure, but with a minority of clades displaying high confidence values: average Bayesian posterior probability was $25.2 \pm 1.95\%$ (average $\pm$ standard deviation) per node. One possibility is that substantial mutational biases in polyguanine markers have resulted in confounding amounts of recurrent mutation and back-mutation, complicating the reconstruction. Alternatively, our data show a limited number of mutations which distinguish cells from one another, which is a known cause of inaccuracy in phylogenetic reconstructions (Feil et al. 2004; Spratt et al. 2004), and which potentially accounts for the relatively low confidence values. This explanation is supported by the observation that many of the phylogeny's clades emanate from a central polytomy, a multifurcating structure which occurs when there are insufficient data for assignment of branching order.

The observation that individual cells demonstrate limited numbers of mutations suggests that the mutation rate of mouse cells *in vivo* is lower than we have previously estimated for cultured cells (Salipante and Horwitz 2006). More markers will be needed to achieve perfect accuracy for all lineages, although the estimated number required depends on assumptions of the models used to represent development. The number of markers available should not be limiting, as there are a high density of polyguanine tracts in mammalian genomes (Salipante and Horwitz 2006). However, in practice, a phylogenetic fate map does not need high confidence assigned to all its clades in order to be useful, as phylogenies can contain isolated credible inferences even if accuracy elsewhere is lacking.

Indeed, our phylogenetic reconstruction does contain valid and potentially valuable lineage inferences. As a functional test of the fate map's accuracy, we included 14 internal controls (designated "a" and "b" followed by the same identification number) from most tissues, where a colony was split into two separate culture vessels after the first six to eight doublings following isolation of the cell from the mouse. It is expected that paired cell lines would have similar genotypes, and correspondingly, should associate most closely on the tree. Encouragingly, 12 of the 14 controls properly grouped together, with each pair sharing a most recent common ancestor that did not give rise to other cells (supplement, phylogeny Newick). (For the other two controls, one pair grouped together in the same clade but were not nearest relatives, whereas cells from the other control did not group together in the same clade). Ten of the paired controls could be considered very credible, with posterior probabilities of 87% or higher, noting that posterior probabilities of $\geq 75\%$ are generally considered trustworthy (Huelsenbeck et al. 2002; Hall 2004). The result argues both for the biological significance of high-probability groupings, and against artifacts resulting from clonal expansion. Further, 10 of the paired control colonies could be differentiated from one another by the presence of identifying mutations, illustrating the degree to which somatic mutations can distinguish lineages.

Excluding the internal controls, the Bayesian tree contained 20 clades with posterior probabilities of 75% or higher, and all of which (like the internal controls) contained only a pair of cells sharing a most recent common ancestor. Eight of those cell pairs were from the same tissue type, whereas the other 12 included cells from two different organs. We were interested in when the most recent common ancestor for each type of grouping had occurred. Thus, we compared the branch length (a measure proportional to the number of divisions cells have undergone [Hall 2004]) from the zygote to the cells' most recent common ancestor and the branch length from that ancestor to the extant cells. The ratio of branch lengths for our internal controls, which are known to share a very recent (artificial) common origin, provide a frame of reference for the comparison (Fig. 2B). The most recent common ancestor of cells from different tissues falls significantly closer to the zygote than the common ancestors of cells from the same tissue. This implies that cell divisions occurring closer to the time of fertilization

frequently give rise to descendants in different tissues, while more recent mitoses have produced cells in the same tissue (Fig. 2C).

Network analysis

Networks can be a useful tool for studying evolutionary relationships when phylogenetic data are complex or conflicting (Huson 1998). In addition to using phylogenetics to examine the relationships between individual cells, we therefore subjected our data to a network analysis in order to search for patterns in how populations of cells from particular tissues developmentally relate to populations from other tissues (Fig. 3). Using a pairwise distance matrix, we calculated each cell's closest neighbors, corresponding to the individuals with the most similar genomes and thus the most similar ancestries. We considered only statistically robust relationships, occurring in 75% or greater of bootstrapped data sets (Felsenstein 1985), which eliminated all but 4103 of the 11,572 original closest neighbor connections. We then compared how frequently cells from a particular tissue grouped together with those from other tissues, and assessed the significance of those comparisons using a χ^2 -test. Overall, this approach provides a measure of how alike the lineage compositions of particular tissues are, which can be displayed graphically (Fig. 3). Nodes represent different tissues, and the lines connecting nodes depict the specific pairwise comparisons on which the network is based. Distance is inversely proportional to the degree of association between populations, so that tissues which are spatially closest in the fate map network represent those, which share the greatest fraction of closely related cells.

The distribution of tissues across the network is not uniform, implying that certain tissues share greater proportions of lineages with other particular tissues. Fat and connective tissue from different sources are most centrally located, and share significant connections among themselves. Other tissues tend to fall more on the periphery. Notably, muscle satellite cells and myoblasts are adjacent to one another, suggesting they are related. Bone marrow stromal cells and cranial vascular endothelial cells appear most removed from the other tissues, implying they derive from a more distantly related cell pool.

Mitotic history of individual tissues

It may be possible to use somatic mutations as a “molecular clock” to estimate the number of mitoses separating an extant cell from the fertilized egg (Salipante and Horwitz 2006; Wasserstrom et al. 2008b). We therefore determined the average number of mutations differentiating cells from the estimated zygote (Fig. 4). Cells harvested from the same tissue tended to exhibit similar numbers of mutations, although differences in mutation density varied by as much as a factor of two across individual tissues ($P = 3 \times 10^{-6}$). At one end of the spectrum, lung fibroblasts appear to have accumulated the fewest mutations since the time of fertilization, whereas bone marrow stromal cells have mutated the most. This finding suggests that cells from those tissues have undergone the fewest number of cell divisions, and the greatest, respectively.

We also investigated the population structure of different tissues. We plotted the distribution of cells carrying various numbers of mutations that distinguish them from the zygote. If all cells in a given population had undergone the same number of divisions, it is expected that such plots would follow a normal distribution with a standard deviation equal to the square root of the mean. (This distribution is effectively equal to the Poisson distribution, with λ equal to the mean.) The data did not fit that model, however (not shown), and we conclude that cells within each population have not undergone a uniform number of mitoses. Modeling studies suggest that our results are consistent either with a single population of cells which has undergone different numbers of divisions, or the presence of two populations of cells within a tissue, in which one population has undergone more mitoses than the other. However, the

distributions for the expected number of mutations per cell under either of these scenarios are nearly identical, making it impossible to distinguish which may be the case biologically.

DISCUSSION

Because of technical limitations, previous fate maps of the mouse have interrogated discrete stages of embryogenesis: some have focused on the events immediately after fertilization (Zernicka-Goetz 2005), others have detailed the period surrounding gastrulation (Beddington 1981; Tam 1989; Lawson et al. 1991; Tam and Behringer 1997), and yet others have examined more terminal development of specific tissues (Eloy-Trinquet et al. 2000; Tremblay and Zaret 2005). Although illuminating, such studies do not provide a cohesive picture of embryogenesis, because they are unable to describe cells' lineage relationships longitudinally across all stages of development. Here, we have examined mouse embryogenesis using the phylogenetic fate mapping approach, permitting interrogation of continuous lineage histories, from the zygote to the adult.

To generate our fate map, we analyzed 298 individual cells cultured from various tissues of a single mouse. Our study utilized a transgenic mouse model, the Immortomouse, which allows for conditional immortalization of cells through selective induction of a SV40 T-antigen oncogene. Because the transgene is not active in vivo, the mouse line is developmentally and reproductively normal, except for variable onset of noncancerous thymic hyperplasia in adulthood (Jat et al. 1991; Vicart et al. 1994). Notably, the young mouse sacrificed in this study showed no signs of that condition. By culturing isolated cells as conditionally immortalized clones, we were able to amplify their DNA sufficiently to allow hundreds of genotyping reactions per cell, and to simultaneously confer a living archive of the primary isolate. We catalogued the unique "fingerprints" of mutation carried by each clone at a collection of 324 polyguanine markers, and used that information to reconstruct lineage relationships.

Phylogenetic algorithms are designed to infer lineages based on polymorphisms accumulated over millions of years of evolution, and because the mitotic mutations we detected here were not nearly as plentiful, it is not surprising that phylogenetic reconstruction of our data did not prove completely robust (Feil et al. 2004; Spratt et al. 2004). We estimated in previous work that approximately 300 markers should be sufficient to infer lineage relationships in mice (Salipante et al. 2008), however, it appears that significantly more will actually be needed. One possibility is that the somatic mutation rate for mouse cells in vivo is less than that we have previously estimated for cells grown in culture (Salipante and Horwitz 2006). Another explanation is that the population structure of the cells has influenced reconstruction fidelity; exponentially growing populations have trees with short, less meaningful branches close to their root (Slatkin and Hudson 1991), an outcome echoed in our reconstruction. Finally, recurrent mutation, manifesting as homoplasy, may further confound accurate reconstructions. Regardless of the cause, interrogation of additional markers in future fate maps should allow for reconstructions with greater fidelity and resolution.

The fate map contains a number of groupings that display high Bayesian posterior probabilities, which our earlier studies suggest can be roughly equated to accuracy (Salipante et al. 2008), and that provide the opportunity for further analysis. The credible inferences of the reconstruction demonstrate that ancestral lineages more frequently give rise to cells of multiple organ systems, whereas relatively recent cell divisions tend to contribute to one tissue type, only (Fig. 2, B and C). These findings imply that lineages mix together at early stages of embryogenesis such that daughter cells can become spatially separated, but that mixing is not significant at later stages. Indeed, those conclusions gain support from experimental observations of mouse embryogenesis made during discrete developmental windows. Previous work utilizing chimerism or cell tagging has also inferred dramatic mixture and migration of

cells during early embryogenesis, even after germ layers arise (Soriano and Jaenisch 1986; Beddington et al. 1989; Condamine et al. 1971; Nesbitt 1971; Saburi et al. 1997). In contrast, terminal cell division in solid organs, as marked by stochastic activation of reporter transgenes, almost always results in contiguous daughter cells (Eloy-Trinquet and Nicolas 2002a, b; Zong et al. 2005). Among the clades we examined, cells that have undergone mixing shared a common lineage over an average of 61.3% of their total cell divisions (Fig. 2B), with a range from 41.5% to 76.2%, implying that at least some mixing events occur even relatively late in development.

Phylogenetic analysis describes the lineage history of individual cells. However, we also considered whether there were patterns to how the collection of lineages which comprise particular tissues may relate developmentally to other types of tissue. To provide a measure of how similar cell populations are between different sources, we quantified how frequently cells from one tissue exhibited closely related relatives in others. The resulting network (Fig. 3) demonstrates that tissues are not uniformly related to one another, suggesting bias in cell lineage allocation. In other words, although any particular tissue is founded by a mixture of different lineages, certain tissues appear to share a more similar composition of those lineages. The simplest explanation is that some tissues derive from similar populations of precursor cells, and thus inherit many of the same lineages. For example, our analysis shows that fibroblasts from muscle, spleen, and lung, have a similar lineage composition, grouping closely together near the center of the network. That finding is consistent with contemporary evidence indicating that fibroblasts are derived from a common, ancient mesenchymal cell pool (Brand-Saberi and Christ 2000). Muscle and spleen fibroblasts appear least similar in lineage composition, suggesting they may derive from separate subpopulations of the primordial fibroblast cell pool. Providing further support for the approach, preadipocytes group closely with fibroblast populations, in agreement with the belief that adipose tissue is developmentally related to connective tissues (Billon et al. 2008). Similarly, cell-labeling techniques have demonstrated that muscle satellite cells and myoblasts share a common embryonic origin in the dermomyotome (Gros et al. 2005), and those two cell populations are accordingly neighbors at one end of the network. (Unfortunately, not enough clones were obtained to make their association statistically robust.) Conversely, bone marrow stromal cells and cranial vascular endothelial cells appear more removed from other tissues, suggesting they arise from significantly different cell pools than those of the other tissues we have studied. Vascular endothelial cells are thought to be derived from locally sequestered mesodermal cells (Coffin and Poole 1988), which our results suggest represent a somewhat different population of lineages than are allocated to fibroblasts. To the best of our knowledge, the embryonic origin of bone marrow stromal cells has not been previously investigated. In summary, the network-based approach recapitulates known information about development, and suggests previously unknown developmental relationships between murine tissues. It will require additional fate maps of multiple mice, perhaps incorporating greater numbers of cells, to determine whether the associations we have observed reflect stochastic underlying processes or if patterns of lineage allocation are reproducible across individuals.

This study presents the first opportunity to examine both early and late events of development in the same individual, and allows for a model of mouse embryogenesis with reference to cell lineages (Fig. 5). Cells labeled at the time of gastrulation do not always exhibit coherent clonal growth (Lawson et al. 1991), suggesting that the interface between the developmental phase marked by cell mixing and that with spatially restricted clonal growth occurs sometime after germ layers are produced. The act of mixing brings cells from different primordial lineage ancestries to the same physical location, resulting in spatially constrained patches of cells from mixed lineages. Those patches are then induced as a unit to differentiate into particular tissues which expand by more cohesive growth of the founding cells. Consequently, although cells of a given organ may be phenotypically indistinguishable in the adult, those populations represent

a mosaic of genotypically distinct early lineages, and retain the genetic evidence of their diverse origins well beyond embryogenesis itself. Given the ability to infer longitudinal lineage histories, it is not possible to neatly classify cells or tissues as derivatives of canonical embryonic domains (such as mesoderm, endoderm, ectoderm, or subdivisions thereof): such embryonic structures are not dedicated clonal populations, but are composed of a similar mixture of many different lineages. Nevertheless, network analysis suggests that particular tissues derive from subtly similar populations of precursor cells, and thus inherit similar proportions of those various ancestral lineages.

The proposed model of mouse development is similar to current understandings of zebrafish (Kimmel et al. 1990; Helde et al. 1994) and chick (Dormann and Weijer 2006; Zamir et al. 2006) embryogenesis, where direct observation of marked cells has revealed early cell mixing that eventually becomes arrested around gastrulation. These findings imply that the model is likely to be generalizable across many species. An intriguing implication of cell mixing, and a possible explanation to why it may be desirable during development of complex organisms, is that it conceivably provides a safeguard against damage and/or stochastic variation during embryogenesis. The ancestral cell pool for any tissue is composed of many different lineages, so that even if there were to be a defective subset of those lineages, sufficient numbers of normal cells could be present for embryogenesis to proceed without conflict. The robustness of mammalian development may, in part, reflect this mechanism: cells may be removed from or added to the pregastrulation mouse embryo without irreversibly perturbing embryogenesis (Power and Tam 1993; Ciemerych et al. 2000), whereas in *C. elegans*, where cell mixing is limited, ablation of single cells is not tolerated (Sulston et al. 1983).

Because the mouse we studied underwent several weeks of development after birth, the data also offer insight into processes occurring postnatally. In cell culture, mutations accrue at a predictable frequency during mitosis, so the number of polymorphisms carried by any cell may be used as a “molecular clock” to determine how many divisions it has undergone from the zygote (Salipante and Horwitz 2006; Wasserstrom et al. 2008b). We therefore calculated the number of mutations distinguishing cells of a particular tissue from the zygote in order to compare the relative number of mitoses cells from various sources had undergone (Fig. 4). We found that individual tissues carry different average numbers of mutations per cell. The data imply that, on average, bone marrow stromal cells have undergone the greatest number of divisions out of the tissues we examined, and conversely, that lung fibroblasts have undergone the fewest mitoses, with about half as many as bone marrow. Although lineage-specific differences in basal mutation rate could potentially skew the number of sequence variants arising in a particular tissue, available experimental evidence does not support this concept (Bielas et al. 2006; Albertson et al. 2009). It is likely that those differences instead reflect dissimilar rates of cell turnover (the loss and replacement of somatic cells) (Rando 2006) in the postnatal mouse. Accordingly, bone marrow, a tissue considered to have a very high turnover rate (Rando 2006), also has the highest mutation density. We found that lung fibroblasts have the lowest number of mutations, and turnover of lung cells, in general, is appropriately considered to be lower than that of many tissues (Pelc 1964; Bowden 1983; Rando 2006), although turnover rates for lung fibroblasts have not been previously reported. Assuming a mutation rate of mouse cells in vivo equal to that determined using fluctuation assays in cultured cells (Boyer et al. 2002), the average number of mutations across all tissues corresponds to about 30 cell divisions (with a range from 22 to 40 mitoses), slightly lower than mathematical estimates for the number of developmental mitoses (Frumkin et al. 2005).

In order to characterize the population structure of tissues, we examined the distribution of mutations across cells from each anatomical source. Contrary to our expectations, none of those data fit the distribution expected if the cells had undergone the same number of mitoses. Thus, cells in each of the individual tissue populations do not demonstrate the same number of past

mitoses, but rather represent a range of previous cell divisions. This finding indicates that cell division does not occur at a uniform rate for all cells in a tissue, but that some have undergone significantly more mitoses than others.

Phylogenetic fate mapping provides a means to simultaneously and longitudinally trace the lineages of many cells from just a single individual—theoretically, every cell in a mouse could be analyzed to generate a comprehensive fate map, analogous to that which has proven so useful in studying *C. elegans* (Sulston et al. 1983). Because the approach is retrospective, it requires no experimental manipulation of the organism, and thus, for the first time, opens up the possibility of constructing human fate maps. To achieve those ambitious goals, however, it is clear that the technology will require refinement in order to improve the amount of useful information it generates. For example, higher resolution reconstructions could be made if it were possible to increase the number of somatic mutations detected, which could be accomplished either by augmenting the number of markers, or by increasing the mutation frequency of the organism, or possibly both. In the near future, another possibility may be to use next-generation sequencing technologies (Eid et al. 2009) to interrogate the genomes of single cells, allowing a complete record of the somatic variation which individualizes them. Regardless, even in its present form, phylogenetic fate mapping offers useful insights into the mechanisms by which organisms develop, and how they maintain themselves postnatally.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank D. Ehlert for help with figures, D. Anderson for helpful discussion, K. F. Benson for help with the mouse, and D. W. McColgin for advice about network visualization. Supported by NIH grants DP1OD003278 and R01DK078340 (to M. S. H.); F30AG030316 (to S. J. S.); NIH T32GM007266, and ARCS Fellowship grants to the University of Washington Medical Scientist Training Program (for S. J. S.).

REFERENCES

- Albertson TM, et al. DNA polymerase epsilon functions as a distinct caretaker of the mouse genome. *Science*. 2009 submitted.
- Beddington RS, Morgernstern J, Land H, Hogan A. An in situ transgenic enzyme marker for the midgestation mouse embryo and the visualization of inner cell mass clones during early organogenesis. *Development* 1989;106:37–46. [PubMed: 2483370]
- Beddington SP. An autoradiographic analysis of the potency of embryonic ectoderm in the 8th day postimplantation mouse embryo. *J. Embryol. Exp. Morphol* 1981;64:87–104. [PubMed: 7310311]
- Bielas JH, Loeb KR, Rubin BP, True LD, Loeb LA. Human cancers express a mutator phenotype. *Proc. Natl. Acad. Sci. USA* 2006;103:18238–18242. [PubMed: 17108085]
- Billon N, Monteiro MC, Dani C. Developmental origin of adipocytes: new insights into a pending question. *Biol. Cell* 2008;100:563–575. [PubMed: 18793119]
- Bowden DH. Cell turnover in the lung. *Am. Rev. Respir. Dis* 1983;128:S46–S48. [PubMed: 6881708]
- Boyer JC, Yamada NA, Roques CN, Hatch SB, Riess K, Farber RA. Sequence dependent instability of mononucleotide microsatellites in cultured mismatch repair proficient and deficient mammalian cells. *Hum. Mol. Genet* 2002;11:707–713. [PubMed: 11912186]
- Brand-Saberi B, Christ B. Evolution and development of distinct cell lineages derived from somites. *Curr. Top Dev. Biol* 2000;48:1–42. [PubMed: 10635456]
- Brownstein MJ, Carpten JD, Smith JR. Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *Biotechniques* 1996;20:1004–1006. 1008–1010. [PubMed: 8780871]

- Ciemerych MA, Mesnard D, Zernicka-Goetz M. Animal and vegetal poles of the mouse egg predict the polarity of the embryonic axis, yet are nonessential for development. *Development* 2000;127:3467–3474. [PubMed: 10903172]
- Clarke JD, Tickle C. Fate maps old and new. *Nat. Cell. Biol* 1999;1:E103–E109. [PubMed: 10559935]
- Coffin JD, Poole TJ. Embryonic vascular development: immunohistochemical identification of the origin and subsequent morphogenesis of the major vessel primordia in quail embryos. *Development* 1988;102:735–748. [PubMed: 3048971]
- Condamine H, Custer RP, Mintz B. Pure-strain and genetically mosaic liver tumors histochemically identified with the—glucuronidase marker in allophenic mice. *Proc. Natl. Acad. Sci. USA* 1971;68:2032–2036. [PubMed: 4109064]
- Dietmaier W, et al. Multiple mutation analyses in single tumor cells with improved whole genome amplification. *Am. J. Pathol* 1999;154:83–95. [PubMed: 9916922]
- Dormann G, Weijer CJ. Imaging of cell migration. *EMBO J* 2006;25:3480–3493. [PubMed: 16900100]
- Drake JW, Charlesworth B, Charlesworth D, Crow JF. Rates of spontaneous mutation. *Genetics* 1998;148:1667–1686. [PubMed: 9560386]
- Eid J, et al. Real-time DNA sequencing from single polymerase molecules. *Science* 2009;323:133–138. [PubMed: 19023044]
- Eloy-Trinquet S, Mathis L, Nicolas JF. Retrospective tracing of the developmental lineage of the mouse myotome. *Curr. Top Dev. Biol* 2000;47:33–80. [PubMed: 10595301]
- Eloy-Trinquet S, Nicolas JF. Cell coherence during production of the presomitic mesoderm and somitogenesis in the mouse embryo. *Development* 2002a;129:3609–3619. [PubMed: 12117811]
- Eloy-Trinquet S, Nicolas JF. Clonal separation and regionalisation during formation of the medial and lateral myotomes in the mouse embryo. *Development* 2002b;129:111–122. [PubMed: 11782405]
- Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J. Bacteriol* 2004;186:1518–1530. [PubMed: 14973027]
- Felsenstein J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol* 1981;17:368–376. [PubMed: 7288891]
- Felsenstein J. Confidence-limits on phylogenies—an approach using the bootstrap. *Evolution* 1985;39:783–791.
- Felsenstein J. Phylogenies from molecular sequences: inference and reliability. *Annu. Rev. Genet* 1988;22:521–565. [PubMed: 3071258]
- Frumkin D, Wasserstrom A, Kaplan S, Feige U, Shapiro E. Genomic variability within an organism exposes its cell lineage tree. *PLoS Comput. Biol* 2005;1:e50. [PubMed: 16261192]
- Gros J, Manceau M, Thome V, Marcelle C. A common somitic origin for embryonic muscle progenitors and satellite cells. *Nature* 2005;435:954–958. [PubMed: 15843802]
- Hall, BG. *Phylogenetic Trees Made Easy*. 2nd Ed.. Sinauer Associates; Sunderland, MA: 2004.
- Helde KA, Wilson ET, Cretekos CJ, Grunwald DJ. Contribution of early cells to the fate map of the zebrafish gastrula. *Science* 1994;265:517–520. [PubMed: 8036493]
- Honig MG, Hume RI. Dil and diO: versatile fluorescent dyes for neuronal labelling and pathway tracing. *Trends Neurosci* 1989;12:333–335. 340–331. [PubMed: 2480673]
- Huelsenbeck JP, Larget B, Miller RE, Ronquist F. Potential applications and pitfalls of Bayesian inference of phylogeny. *Syst. Biol* 2002;51:673–688. [PubMed: 12396583]
- Hughes S, Arneson N, Done S, Squire J. The use of whole genome amplification in the study of human disease. *Prog. Biophys. Mol. Biol* 2005;88:173–189. [PubMed: 15561304]
- Huson DH. SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* 1998;14:68–73. [PubMed: 9520503]
- Jackson AL, Loeb LA. The contribution of endogenous sources of DNA damage to the multiple mutations in cancer. *Mutat. Res* 2001;477:7–21. [PubMed: 11376682]
- Jat PS, et al. Direct derivation of conditionally immortal cell lines from an H-2Kb-tsA58 transgenic mouse. *Proc. Natl. Acad. Sci. USA* 1991;88:5096–5100. [PubMed: 1711218]
- Kimmel CB, Warga RM, Schilling TF. Origin and organization of the zebrafish fate map. *Development* 1990;108:581–594. [PubMed: 2387237]

- Klein CA, Schmidt-Kittler O, Schardt JA, Pantel K, Speicher MR, Riethmuller G. Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. *Proc Natl. Acad. Sci. USA* 1999;96:4494–4499. [PubMed: 10200290]
- Lawson KA, Meneses JJ, Pedersen RA. Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. *Development* 1991;113:891–911. [PubMed: 1821858]
- Livet J, et al. Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* 2007;450:56–62. [PubMed: 17972876]
- Nesbitt MN. X chromosome inactivation mosaicism in the mouse. *Dev. Biol* 1971;26:252–263. [PubMed: 5158534]
- Pelc SR. Labelling of DNA and cell division in so called non-dividing tissues. *J. Cell. Biol* 1964;22:21–28. [PubMed: 14195612]
- Power MA, Tam PP. Onset of gastrulation, morphogenesis and somitogenesis in mouse embryos displaying compensatory growth. *Anat. Embryol. (Berlin)* 1993;187:493–504. [PubMed: 8342794]
- Rando TA. Stem cells, ageing and the quest for immortality. *Nature* 2006;441:1080–1086. [PubMed: 16810243]
- Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 2003;19:1572–1574. [PubMed: 12912839]
- Saburi S, Azuma S, Sato E, Toyoda Y, Tachi C. Developmental fate of single embryonic stem cells microinjected into 8-cell-stage mouse embryos. *Differentiation* 1997;62:1–11. [PubMed: 9373942]
- Salipante, SJ.; Horwitz, M. A phylogenetic approach to mapping cell fate.. In: Schatten, GP., editor. *Current Topics in Developmental Biology*. Academic Press; 2007. p. 157-184.
- Salipante SJ, Horwitz MS. Phylogenetic fate mapping. *Proc. Natl. Acad. Sci. USA* 2006;103:5448–5453. [PubMed: 16569691]
- Salipante SJ, Thompson JM, Horwitz MS. Phylogenetic fate mapping: theoretical and experimental studies applied to the development of mouse fibroblasts. *Genetics* 2008;178:967–977. [PubMed: 18245843]
- Slatkin M, Hudson RR. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 1991;129:555–562. [PubMed: 1743491]
- Soriano P, Jaenisch R. Retroviruses as probes for mammalian development: allocation of cells to the somatic and germ cell lineages. *Cell* 1986;46:19–29. [PubMed: 3013418]
- Spratt BG, Hanage WP, Li B, Aanensen DM, Feil EJ. Displaying the relatedness among isolates of bacterial species—the eBURST approach. *FEMS Microbiol. Lett* 2004;241:129–134. [PubMed: 15598523]
- Stern CD, Fraser SE. Tracing the lineage of tracing cell lineages. *Nat. Cell. Biol* 2001;3:E216–E218. [PubMed: 11533679]
- Sulston JE, Schierenberg E, White JG, Thomson JN. The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol* 1983;100:64–119. [PubMed: 6684600]
- Tam PP. Regionalisation of the mouse embryonic ectoderm: allocation of prospective ectodermal tissues during gastrulation. *Development* 1989;107:55–67. [PubMed: 2627894]
- Tam PP, Behringer RR. Mouse gastrulation: the formation of a mammalian body plan. *Mech. Dev* 1997;68:3–25. [PubMed: 9431800]
- Thompson JM, Salipante SJ. PeakSeeker: a program for interpreting genotypes of mononucleotide repeats. *BMC Res. Notes* 2009;2:17. [PubMed: 19192296]
- Tremblay KD, Zaret KS. Distinct populations of endoderm cells converge to generate the embryonic liver bud and ventral foregut tissues. *Dev. Biol* 2005;280:87–99. [PubMed: 15766750]
- Vicart P, et al. Immortalization of multiple cell types from transgenic mice using a transgene containing the vimentin promoter and a conditional oncogene. *Exp. Cell Res* 1994;214:35–45. [PubMed: 8082738]
- Wasserstrom A, et al. Reconstruction of cell lineage trees in mice. *PLoS One* 2008a;3:e1939. [PubMed: 18398465]
- Wasserstrom A, et al. Estimating cell depth from somatic mutations. *PLoS Comput. Biol* 2008b; 4:e1000058. [PubMed: 18404205]

- Zamir EA, Czirók A, Cui C, Little CD, Rongish BJ. Mesodermal cell displacements during avian gastrulation are due to both individual cell-autonomous and convective tissue movements. *Proc. Natl. Acad. Sci. USA* 2006;103:19806–19811. [PubMed: 17179040]
- Zernicka-Goetz M. Cleavage pattern and emerging asymmetry of the mouse embryo. *Nat. Rev. Mol. Cell Biol* 2005;6:919–928. [PubMed: 16341078]
- Zong H, Espinosa JS, Su HH, Muzumdar MD, Luo L. Mosaic analysis with double markers in mice. *Cell* 2005;121:479–492. [PubMed: 15882628]

Fig. 1.

Examples of and statistics for somatic variation in polyguanine markers. (A) Electropherograms of a polyclonal mixture of cells used to estimate the genotype of the mouse zygote aligned with those of various somatic cells. X-axis indicates product length (bp), Y-axis represents signal intensity. Allele lengths are indicated, with mutations from the zygote noted. (B) Histogram for the number of markers that demonstrate different numbers of distinguishable mutant genotypes, across all cells. (C) Histogram for the quantity of cells displaying particular numbers mutations from the zygote.

Fig. 2.

Lineage reconstruction. (A) The Bayesian phylogenetic fate map is displayed as a circular phylogram rooted at the estimated zygote. Scale bar reflects the number of changes per allele along branches. Each terminal circle represents a single cell, color coding indicates the cell's origin. (B) Fraction of the branch length from the zygote to the most recent common ancestor of cells from the same tissue or two different tissues. Internal controls, which are known to share a very recent common ancestor, are shown for comparison. Error bars indicate standard error of the mean. * $P < 0.05$, ** $P < 0.001$. (C) Depiction of results from B. Recent mitoses, with the greatest distance from the common ancestor ("A," in circle) to the zygote, tend to produce

cells from the same tissue (top). Older common ancestors, falling closer to the zygote, tend to produce cells of different tissues (bottom).

Fig. 3.

Network-based fate map of mouse tissues. The degree of similarity between tissues' founding lineages is depicted. Nodes represent various tissues, and the lines connecting them depict the specific pairwise comparisons used to construct the network. The distance between nodes (or along lines) is inversely proportional to the degree of association between two tissues. Black lines indicate highly significant ($P < 0.05$) relationships, gray lines represent suggestive relationships ($P < 0.25$), and less significant relationships are not shown. It is possible for connections to extend unidirectionally from one node, thus, arrowheads represent the polarity of association.

Fig. 4. Mitotic distance from the zygote to tissues. The average number of mutations differentiating cells from the estimated zygote are displayed for each tissue. Error bars indicate standard error of the mean.

Fig. 5. Model of embryogenesis. Fertilization establishes the genotype of the diploid zygote. As mitosis occurs, lineages become distinguishable from one another based on their pattern of somatic mutations. Cell divisions during cleavage are thought to result in contiguous daughter cells (Zernicka-Goetz 2005). Extensive cell mixing occurs before gastrulation, homogenizing lineages. Cell motility becomes restricted sometime after gastrulation. During organogenesis, patches of contiguous cells, derived from dissimilar lineages, are induced to form particular tissues. Tissues increase in size through relatively coherent clonal growth of various lineages in the induced patches.