



## OPEN ACCESS

## EDITED BY

Alexandre V. Morozov,  
Rutgers, The State University of New Jersey,  
United States

## REVIEWED BY

Toni Gossmann,  
Technical University Dortmund, Germany

## \*CORRESPONDENCE

Seong Gyu Kwon,  
sgkwon@gnu.ac.kr

RECEIVED 06 December 2025

REVISED 27 December 2025

ACCEPTED 29 December 2025

PUBLISHED 09 January 2026

## CITATION

Sajjad M and Kwon SG (2026) Reconstructing developmental lineages: a retrospective approach using somatic mutations and variant allele frequency.

*Front. Genet.* 16:1761810.

doi: 10.3389/fgene.2025.1761810

## COPYRIGHT

© 2026 Sajjad and Kwon. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Reconstructing developmental lineages: a retrospective approach using somatic mutations and variant allele frequency

Mahnoor Sajjad and Seong Gyu Kwon\*

Department of Anatomy and Convergence Medical Science, College of Medicine, Institute of Medical Science, Gyeongsang National University, Jinju, Republic of Korea

Somatic mutations accumulate during the first zygotic division and continue throughout an organism's lifespan. The characteristics and frequency of these mutations are contingent on developmental timing and tissue type, giving rise to somatic mosaicism, defined as the presence of unique genomic alterations across different cells. They serve as endogenous cellular barcodes, enabling detailed reconstruction of cell lineages and clonal dynamics. Although lineage tracing techniques have advanced from early microscopic observation and dye staining to the introduction of artificial barcodes via gene editing, owing to ethical considerations, such genetic manipulations in human developmental research are unavailable. Therefore, spontaneously arising somatic mutations are the most suitable strategy for tracing human lineages. Current approaches can be broadly categorized into two strategies: (i) high-resolution methods, including single-cell clonal expansion or laser-capture microdissection, which construct precise phylogenetic trees based on shared mutation profiles; and (ii) bulk sequencing methods, which infer lineage proximity by comparing variant allele frequencies across samples. As more lineage-tracing studies are being conducted focusing on a wider variety of organs, the integration of such data will make it possible to discover the general principles governing human development. This review highlights how the concept of somatic mutations has been applied across diverse biological contexts and discusses the insights and common principles that can be drawn from these findings.

## KEYWORDS

embryogenesis, lineage tracing, lineage tree, somatic mutation, variant allele frequency (VAF), whole-genome sequencing (WGS)

## 1 Introduction

During somatic cell division, errors that occur during DNA replication can result in permanent changes in the DNA sequence, known as somatic mutations ([Lodato and Vijg, 2022](#)). These somatic mutations can be induced by both internal and external factors, including replication errors and environmental effects. These mutations are permanently "scared" in the DNA of the daughter cells if not corrected ([Loeb and Cheng, 1990](#)).

These somatic mutations accumulate continuously from the first division of the zygote throughout the growth of the organism ([Mohiuddin et al., 2022](#)), and their characteristics

and frequency vary depending on developmental timing and tissue type in terms of internal factors (Kim et al., 2022).

Consequently, individual cells in the same organism have distinct somatic mutations, making cell populations genetically heterogeneous. This phenomenon is referred to as somatic mosaicism, which describes the presence of unique genomic alterations in different tissues or cells depending on the timing and location of somatic mutations that occur after zygote formation (Freed et al., 2014).

Somatic mosaicism indicates that mutations at the single-cell level can be distributed across multiple cells and tissues within an organism. Recent studies have shown that somatic mutations can occur in different ways and in different tissues at various stages of development (De, 2011).

Somatic mutations include various genetic alterations that occur in cells after fertilization. The types of somatic mutations include single-nucleotide variants (SNVs), small insertions and deletions (indels), and large structural variations, such as deletions, duplications, and translocations, as well as copy number variations (Pleasance et al., 2010).

In clinical research, particularly in cancer genetics, somatic mutations are often classified based on their functional roles. For example, somatic variants can be categorized as oncogenic, likely oncogenic, variants of uncertain significance, likely benign, or benign, depending on their potential to drive disease processes, such as tumorigenesis (Horak et al., 2022).

Somatic mutations are the key drivers of the initiation and progression of many cancers. Somatic mutations in EGFR and KRAS are key drivers of lung cancer. EGFR mutations are associated with a favorable response to EGFR tyrosine kinase inhibitors (TKIs) and improve overall survival. In contrast, KRAS mutations are associated with reduced responsiveness to EGFR-TKIs and predict shorter survival in patients with advanced lung adenocarcinoma (Johnson et al., 2013; Wang et al., 2024). In breast cancer, driver mutations frequently occur in genes such as PIK3CA and TP53. PIK3CA mutations are most common in estrogen receptor (ER)-positive breast cancers, whereas TP53 mutations are predominant in ER-negative cases. TP53 mutations are associated with poorer prognosis in ER-positive metastatic breast cancer, whereas in ER-negative cases, TP53 mutations may have a protective effect (Kim et al., 2017; Rajendran and Deng, 2017). For colorectal cancer, the most recurrent driver mutations have been found in APC, TP53, and KRAS. These mutations contribute to clonal expansion and chemoresistance of cancer cells. Concurrent KRAS and TP53 mutations are strongly associated with poor response to standard chemotherapy, increased risk of recurrence and metastasis, and worse overall prognosis (Matas et al., 2022; Tang and Fan, 2024). Mutations in BRCA1/2 and TP53 are important for ovarian cancer. BRCA1/2 mutations are associated with increased sensitivity to platinum-based chemotherapy and PARP inhibitors, resulting in improved survival. However, TP53 accumulation can modify the prognostic impact of BRCA1 mutations, particularly in high-grade serous ovarian carcinomas (Pennington et al., 2014; Rzepecka et al., 2017).

Recent studies revealed that somatic mutations are also present in normal tissues, with their accumulation beginning from the first

cell division after conception (Rahal et al., 2024). The number of accumulated somatic mutations varies significantly among different tissue types and species (Cagan et al., 2022). For example, mutation rates are higher in tissues such as the colon and skin than in germline cells or the brain (Werner and Sottoriva, 2018).

The accumulation of somatic mutations is influenced by various factors, including environmental exposure (such as tobacco smoke or UV light), cell division rates, and DNA repair efficiency, leading to individual- and tissue-specific differences in mutation burden (Ren et al., 2022). The pattern of somatic mutation accumulation is known as a mutational signature. A comprehensive analysis of mutational signatures in normal tissues has shown that mutational processes similar to those found in cancer cells are also found in normal cells (Yaacov et al., 2023; Boysen et al., 2025).

In addition to analyzing somatic mutations in diseases or cancers, somatic mutations can also be used as natural barcodes for retrospective cellular lineage tracing (Behjati et al., 2014). Since somatic mutations occur at random sites across the genome, they act as distinct cellular identifiers, enabling the detailed tracing of cell lineages and clonal dynamics (Park et al., 2021). Multiple studies have applied lineage tracing based on somatic mutations in diverse tissues to discover lineage relationships between cells and tissues from various organs (Bae et al., 2018; Coorens et al., 2021a; Park et al., 2021; Spencer Chapman et al., 2021).

Retrospective lineage tracing using somatic mutations fundamentally relies on the fact that starting from the zygote, mutations accumulate uniquely as cells divide. As cells continuously divide throughout their lifetime, they accumulate genetic mutations, enabling the inference of their division history. In other words, the presence or absence of shared somatic mutations among cells allows for the complete reconstruction of their developmental trajectories (Choi et al., 2023).

As research on lineage tracing using somatic mutations has increased, there is a growing need for techniques to construct phylogenetic trees based on somatic mutations identified in various tissues from a single organism. However, differences in variant filtering criteria and tree construction algorithms among the studies have been identified as limitations. A recent study presented detailed guidelines for building phylogenetic trees using somatic mutations, suggesting that these could serve as standards for lineage-tracing research (Coorens et al., 2024).

The variant allele frequency (VAF) concept has been applied to somatic mutations in lineage-tracing studies. The VAF derived from bulk sequencing data, where cells of multiple origins are mixed, reflects the relative prevalence of a given somatic mutation within the sample (Dou et al., 2018). In other words, VAF serves as an indicator of the contribution of cells with a particular somatic mutation to the cell population in the bulk tissue.

In this review, we analyze studies that utilized somatic mutations for developmental lineage tracing. We highlight how this concept has been applied across diverse biological contexts and discuss the insights and common principles that can be drawn from these findings.

## 2 Methods for developmental lineage tracing

### 2.1 Early studies

One of the earliest documented studies on developmental lineage tracing was conducted by Charles Otis Whitman in the late 19th century. Whitman observed and mapped the fate of each cell during leech embryonic development (Whitman, 1878). His work suggests that the fate of cells in the developmental stage is determined in the early cleavage stage and is not a stochastic process as previously thought.

Sulston and his colleagues traced the complete lineage of every cell in *Caenorhabditis elegans* from fertilization to adulthood (Sulston et al., 1983). This was a milestone study in developmental biology, and the authors reported when and how the fate of cells during early development was determined. Additionally, they found that the cells disappeared naturally, proposing the idea of programmed cell death.

Given the limitations of microscopic observation of cells, studies have been conducted to analyze cell fate using vital dyes. The dye must effectively stain the target cells without being harmful to the cells. This method has been used in studies that traced the development of *Xenopus* up to the 32-cell stage (Vogt, 1929) and in a study that determined the fate map of the zebrafish neural plate (Woo and Fraser, 1995). However, the use of dyes has disadvantages: they can leak into adjacent cells, and their concentration is diluted with each cell division, leading to a decrease in accuracy.

### 2.2 Implications of genetic tools

The rapid advancement in genetic tools since the 1990s has improved approaches to lineage tracing, enabling more precise tracking of developmental processes than in previous studies. The Cre-Lox-based technique is the most widely used for modern genetic lineage tracing. The Cre-loxP system regulates the tissue-specific activation or inhibition of gene function, enabling progenitor cells to produce descendant cells marked by reporter genes that serve as permanent and inheritable genetic barcodes (Wang et al., 2023). Several studies have been conducted on lineage tracing across different tissues using this method. In the lung, it has been revealed how club cells regenerate ciliated cells following airway injury (Rawlins et al., 2009). Additionally, this approach has revealed how epicardial progenitors contribute to cardiomyocyte formation during heart development (Zhou et al., 2008). Similar methodologies have been applied to investigate the origin of neural cells (Adameyko and Lallemand, 2010). Taken together, Cre-lox-based lineage tracing is a powerful tool capable of tracking cell fate across a broad timeline ranging from early developmental stages to specific cell populations at later developmental time points.

Similar to the Cre-loxP system, CRISPR-Cas9-based lineage tracing introduces mutations that act as barcodes for the reconstruction of cellular phylogenies. This approach has been applied to track developmental processes in animals. For instance, in zebrafish, CRISPR-Cas9 has been used to create mutations in early embryos, allowing researchers to build

comprehensive lineage trees that reveal the relationships between different cell types as an organism develops. Furthermore, in a study on pancreatic cancer metastasis, researchers used the CRISPR-Cas9 tool macsGESTALT to create unique genetic markers for individual cancer cells in a mouse model. This enabled them to track how thousands of these single cells spread throughout the body to form new tumors in different locations (Simeonov et al., 2021).

Taken together, genetic tools have significant advantages for lineage tracing by providing permanent and heritable markers, unlike the traditional methods of observing embryos under a microscope or using dyes.

### 2.3 Retrospective lineage tracing using somatic mutations

Despite the advantages of genetic tools, their application for prospective lineage tracing in humans is not feasible. There are ethical concerns about intentionally introducing permanent genetic variations into human participants, even for research purposes (Almeida and Ranisch, 2022). Therefore, alternative methods that do not raise ethical concerns are required for studying lineage tracing in humans.

As described above in the Introduction, somatic mutations occur naturally during cell division, and each mutation acts as a specific DNA barcode that accumulates in different patterns in each cell. The core principle is that as an embryo develops, distinct barcode patterns are introduced into the genome of each daughter cell in the form of mutations. These mutations accumulate with each subsequent cell division. Consequently, by the time development is complete, the DNA of cells in the adult organism will be 'scared' with a unique and diverse pattern of these mutational barcodes, reflecting their lineage history (Figure 1).

Since somatic mutations are present in every cell in a unique pattern, single-cell analysis enables the most accurate reconstruction of cell lineage tracings. Genomic DNA in single cells can be amplified for whole-genome sequencing (WGS) using whole-genome amplification (WGA). Although WGA is a powerful tool for discovering somatic mutations in a single cell, it creates a large number of errors, including false mutations (Dou et al., 2018) and biased amplification of two alleles (Huang et al., 2015).

Cell proliferation through cell division is a form of DNA amplification. The key difference between WGA techniques is that one occurs within the cell, whereas the other is an artificial process outside the cell. The accuracy of intracellular DNA replication with repair mechanisms is considerably higher than that of WGA (more than 1,000 times) (Youk et al., 2021). Therefore, to accurately identify somatic mutations in a specific cell, clonal expansion of single cells can be used to obtain sufficient DNA for analysis. However, not all cells, especially differentiated cells, are difficult to culture *in vitro* to acquire sufficient amounts of DNA (Bukowy-Bieryło, 2021). Therefore, the cell types suitable for lineage tracing using this method are limited. Various alternative methods have been developed to extend the application of somatic mutation-based lineage tracing to a broad range of tissues.

Certain stem cells grow clonally in specific compartments (niches) *in vivo* (Simons and Clevers, 2011; Blanpain and Simons, 2013). This process is usually observed in the intestinal

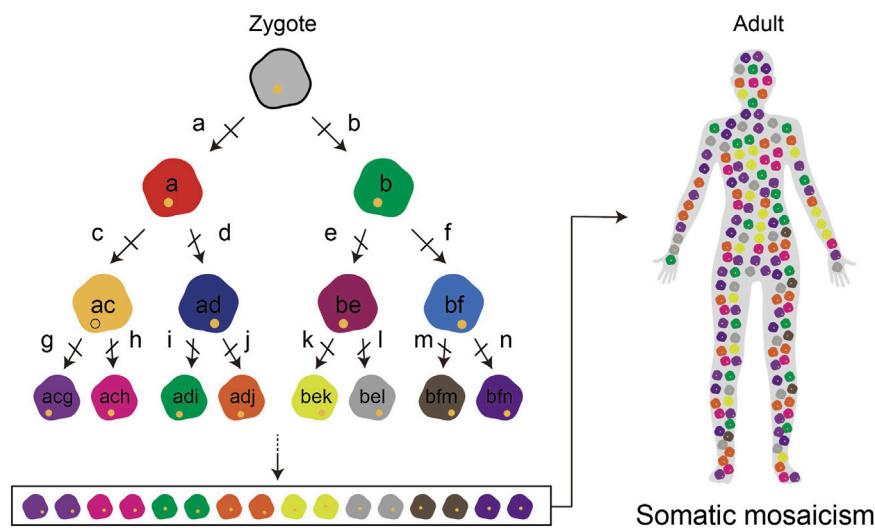


FIGURE 1

Somatic mutations result in somatic mosaicism in adults. Somatic mutations, occurring at each cell division, are represented as 'a' through 'n.' The final combination of mutations provides each cell with a unique set of characteristics. Consequently, the distinct properties of adult tissues show a mix of different cells, referred to as somatic mosaicism.

crypts (Snippert et al., 2010). A recent study quantitatively confirmed the occurrence of clonal proliferation. The clonal expansion of cells has been observed primarily in the glands within the digestive tract (Moore et al., 2021). Cell populations formed by the clonal growth of a specific stem cell can theoretically be considered analogous to populations generated through single-cell clonal expansion, as described earlier. Once clonal populations are identified under a microscope, a specialized technique called laser-capture microdissection (LCM) can be used to physically isolate the niche region for WGS analysis (Ellis et al., 2021). Using this method, the quantity and quality of the DNA extracted for WGS are poorer than those with single-cell clonal expansion but still demonstrate higher levels of quantity and quality than in WGA (Youk et al., 2021).

To overcome the limitations of WGA, a recent method called primary template-directed amplification (PTA) enables the uniform amplification of DNA from single cells or small numbers of cells, allowing for the accurate detection of somatic mutations (Gonzalez-Pena et al., 2021). Therefore, PTA can be used to detect somatic mutations in tissues where single-cell clonal expansion is impossible and the *in vivo* niche is difficult to identify, such as the brain (Kalef-Ezra et al., 2024).

Ultimately, various approaches for inferring developmental processes in normal cells share the common goal of accurately detecting somatic mutations embedded in the DNA for retrospective lineage tracing.

### 3 Lineage tracing studies in humans

#### 3.1 Building direct lineages with clonally expanded cell populations

Most lineage-tracing studies have been conducted by constructing phylogenetic trees that focus on specific organs. A

representative example is the collection of skin fibroblasts from donated cadavers, followed by single-cell clonal expansion and phylogenetic reconstruction (Park et al., 2021). In terms of scale, the cited study by Park et al. surpasses previous studies by the number of people included (five cadavers) and by collecting more samples (more than 300 samples). In their study, samples were collected from multiple anatomical sites on the body. Based on the somatic mutation patterns identified in each sample, phylogenetic trees were reconstructed to relate the anatomical location to lineage relationships. Additionally, they used simulations to determine the cell division stage at which the epiblast and trophoblast lineages were segregated from the zygote. The most significant finding of this study is that identifying such a cell division stage demonstrates the possible reason why the two daughter cells arising from the first division contribute to the whole body in different proportions, which is called "asymmetric distribution." Subsequently, this hypothesis was experimentally verified by detailed observations of the actual division of embryos (Junyent et al., 2024).

The greater the number and diversity of samples analyzed, the more precisely the later cell division stages can be inferred. Therefore, this study provides an important benchmark for future lineage-tracing analyses.

Concurrently, phylogenetic trees were reconstructed from various tissues in other studies. Using LCM, this research enabled lineage tracing through somatic mutation analysis in a diverse range of cells, overcoming the prior limitation of single-cell clonal expansion techniques, which were restricted to specific cell types (Coorens et al., 2021a). The study also confirmed the phenomenon of asymmetric distribution in cell lineages and observed when the fates of the trophectoderm and inner cell mass were determined. These findings provide crucial clues for tracking embryonic development in humans.

Additionally, some studies have analyzed the lineage of specific organs or tissues. Lodato et al. examined somatic mutations using single-cell sequencing of neurons from three normal brains (Lodato

et al., 2015). They constructed a phylogenetic tree and identified lineage distances between different brain regions. Bae et al. analyzed somatic mutations in single-cell clonal expansion samples from the forebrains of three fetuses (Bae et al., 2018). They found 200–400 SNVs in each cell and constructed small lineage trees based on the sharedness of the mutations. It is noteworthy that single-cell clonal expansion technology has enabled the discovery of more accurate mutations than single-cell sequencing used in previous brain studies.

Lineage tracing techniques have also been applied to the study of hematopoiesis. Hematopoietic stem cells (HSCs) and progenitor cells (HPCs) were isolated from the bone marrow. Following single-cell clonal expansion, somatic mutations were analyzed to construct a phylogenetic tree. This tree was subsequently used to define the lineage relationship between the 2 cell types and estimate the effective population size contributing to blood cell production (Lee-Six et al., 2018). Similarly, other studies constructed lineage trees from HSCs and HPCs and analyzed the signatures of accumulated somatic mutations to investigate their association with leukemia (Osorio et al., 2018). The group led by Peter J. Campbell has conducted extensive research using lineage tracing in hematopoietic cells to investigate a wide range of associated biological phenomena. They collected a substantially larger number of cells from fetal hematopoietic organs than previous studies. A phylogenetic tree constructed from these cells was used to estimate the divergence time between the embryonic and extraembryonic lineages (Spencer Chapman et al., 2021). This finding is consistent with results previously observed in fibroblasts (Park et al., 2021). The similarity in results, even when constructing independent lineage trees from different tissues, indicates that most cells share a common developmental fate during early cell division. Further investigations involved the collection of hematopoietic cells from donors of various ages to construct lineage trees (Mitchell et al., 2022). These trees were used to evaluate changes in clonal structure and emergence timing of driver mutations. A key finding was that the hematopoietic clonal diversity diminished with age. Beyond lineage tracing, the same research group examined how clonal diversity is altered in patients after hematopoietic cell transplantation (Spencer Chapman et al., 2024). Finally, the study was extended to mice by performing lineage tracing using a similar method and cell type (Kapadia et al., 2025). This comparative analysis with human data demonstrates the potential for future applications in other species.

Studies of lineage tracing in other organs are limited. For example, Tim et al. used LCM to sample tissues from the placenta (Coorens et al., 2021b) and gastric epithelium (Coorens et al., 2025) and performed lineage tracing on a small scale. Technological advancements in single-cell manipulation are expected to enable the application of lineage tracing to a wide variety of organs.

### 3.2 Inferring lineage relationships from bulk sequencing samples

Using single-cell clonal expansion and LCM technology, we can estimate lineage relationships at the single-cell level or at a similar resolution. However, when single-cell culture is

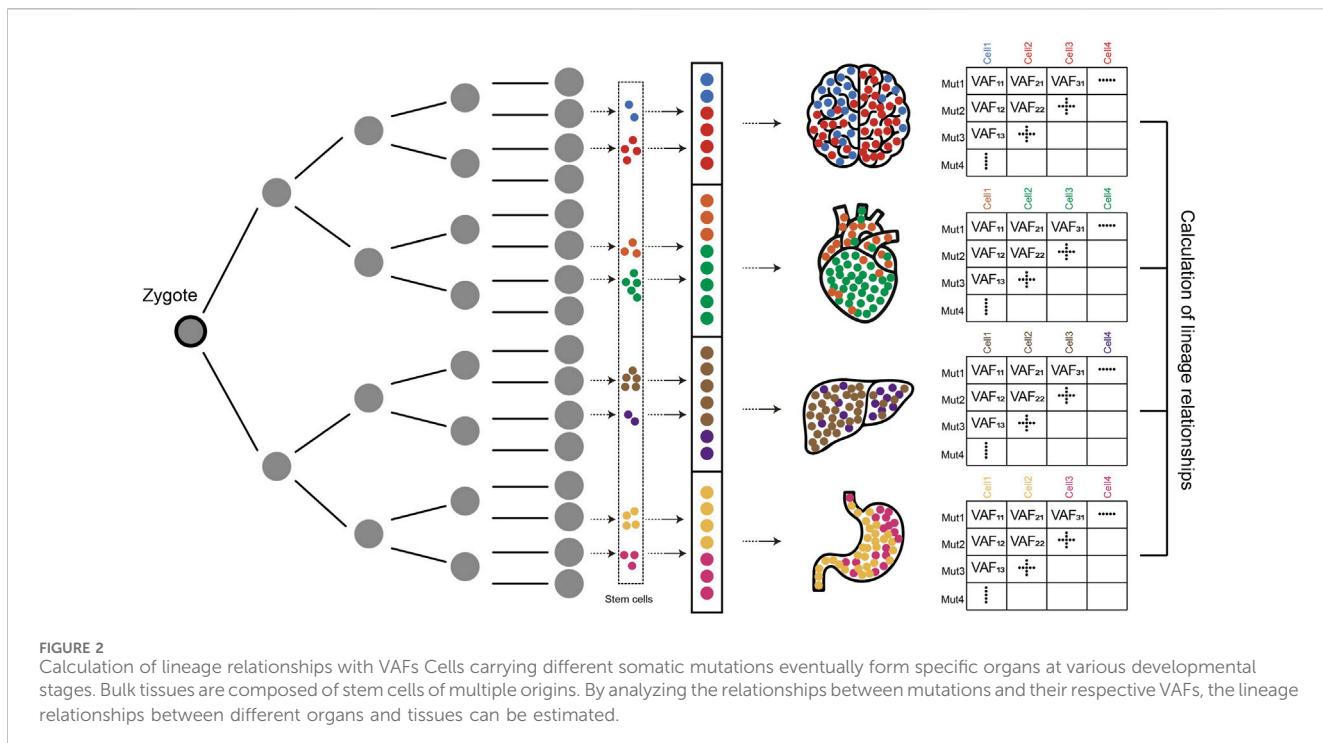
impossible or tissue characteristics prevent LCM from isolating clonal populations, indirect lineage tracing is possible by analyzing the bulk tissue. Unlike single-cell clonal expansion samples, in bulk samples, somatic mutations exist not in a presence or absence state but as values of varying degrees, and the metric representing these values is the VAF (Moeller et al., 2023). Bulk samples are collections of single cells of multiple origins containing diverse combinations of somatic mutations, and at some point during development, stem cells expand to predominantly form specific tissues or organs. Therefore, the lineage relationship between the bulk samples was estimated by comparing the similarity of VAFs of the somatic mutations (Figure 2).

One study analyzed somatic mutations in blood using bulk WGS, estimating the developmental stage at which each mutation occurred based on its VAF. This study revealed the unequal contribution of early embryonic cells to adult somatic tissues (Ju et al., 2017). Notably, this finding supports a conclusion similar to that reported by Park et al. (2021) using a single-cell clonal expansion analysis.

In 2021, two similar studies identified somatic mutations in multiple organ tissues through bulk WGS and target sequencing, and estimated developmental lineage tracing (Bizzotto et al., 2021; Fasching et al., 2021). A consistent finding from both studies was the asymmetric contribution of early progenitors to extraembryonic tissues, three germ layers, and organs.

Unlike the studies described above, a recent study provided detailed analyses of the developmental processes of a specific organ. Bulk WGS was conducted on tissue samples from multiple regions of the left and right brain to identify somatic mutations (Breuss et al., 2022). Pearson's correlation coefficients were calculated for these mutations using their VAFs across the collected samples, followed by clustering analysis. The results demonstrated location-specific grouping, where mutations from distinct brain regions were clustered together. An important finding of this study is that the migration of brain precursor cells across the anterior-posterior and ventral-dorsal axes is constrained after a specific time during development, which was inferred from the distinct spatial distribution of somatic mutations. Furthermore, we identified somatic mutations in single nuclei via amplicon sequencing to confirm the lineage relationships. However, it still has limitations in that it does not demonstrate the same level of accuracy as techniques such as clonal expansion or LCM.

Bulk WGS enables the identification of somatic mutations and the inference of lineage relationships regardless of the tissue type. However, this approach has a limitation in that mutations acquired later in development often exhibit low VAFs, making their detection challenging (Menon and Brash, 2023). Low-VAF mutations are supported by few reads that contain mutations, and it is difficult to distinguish true variants from technical artifacts without harsh filtering methods (Huang and Lee, 2021). To overcome these limitations, technologies for detecting low-VAF variants, such as duplex sequencing (Kennedy et al., 2014) and NanoSeq (Abascal et al., 2021), are being developed. Duplex sequencing tags both DNA strands to generate a consensus. By validating mutations on both strands, it eliminates artifacts and enables the detection of ultra-low-VAF variants. NanoSeq improves duplex consensus by minimizing



library preparation artifacts. With an ultra-low error rate (<5 per billion), it detects even single-molecule mutations with high specificity.

## 4 Lineage tracing studies in other species

Somatic mutation-based lineage tracing is mostly used in human studies as it circumvents ethical concerns. Concurrently, this methodology can also be applied to animal models to provide a comparative framework for understanding developmental processes in humans.

The first organismal lineage-tracing study at the single-cell level was conducted using mouse-derived organoids (Behjati et al., 2014). Although the scale of this study was small, it revealed that the two daughter cells from the initial zygotic division were unequally distributed in the adult tissue, a finding consistent with subsequent studies.

The application of single-cell clonal expansion to lineage tracing has been extended to pigs (Kwon et al., 2024). Parallel to the findings of human studies, this study demonstrated an asymmetric contribution from early developmental cells in pigs.

In another study, bulk WGS was applied across multiple mouse organs to investigate lineage relationships by analyzing somatic mutations and their VAFs (Uchimura et al., 2022). Their study demonstrated that for VAF-based lineage tracing, it is crucial to obtain accurate VAFs through high-coverage WGS, and VAF values must be acquired from multiple tissue sources.

Lineage tracing in non-human animals can facilitate comparative analysis of mammalian development and validate the utility of experimental disease models for human pathologies.

## 5 Discussion

Somatic mutations occur during development and leave permanent genomic barcodes across descendant cells. This gives rise to somatic mosaicism and intra-individual genetic heterogeneity in normal tissues. Therefore, mutational fingerprints act as permanent records of an organism's developmental history.

Lineage tracing methods using somatic mutations are employed to estimate human development when invasive techniques such as gene editing cannot be used. The application of WGS to single-cell clonal expansion samples has enabled the reconstruction of lineage trees at single-cell resolution, offering detailed insights into the developmental history of organisms and specific tissues. A prominent finding from these analyses is the unequal contribution of cells from the early developmental stages to adult tissue composition.

The reconstruction of accurate lineage trees not only explains the developmental processes of normal tissues but also provides crucial insights into pathogenesis, enabling the inference of when disease-initiating mutations occur. This methodology has been extensively applied to hematopoietic cells to determine the developmental timing of various mutations associated with hematological malignancies. This approach has significant potential for identifying the developmental origins of cancers and genetic disorders across diverse tissue types.

Bulk WGS is not suitable for reconstructing formal lineage trees but has broad applicability, enabling the estimation of lineage relationships from nearly all tissue types.

This review surveyed major studies on lineage tracing in normal tissues using somatic mutations, and we observed that research focused on specific organs is not yet widely diversified. As more

lineage-tracing studies are being conducted focusing on various organs, the integration of such data will make it possible to discover the general principles of human development.

In conclusion, analysis of lineage relationships at the cellular and tissue levels using somatic mutations offers a valuable solution when other lineage-tracing methods are not applicable. Furthermore, these approaches are versatile and are capable of applying somatic mutations to binary (e.g., mutation presence/absence) and continuous (e.g., VAFs) data.

## Author contributions

MS: Data curation, Visualization, Writing – original draft. SGK: Writing – review and editing.

## Funding

The author(s) declared that financial support was received for this work and/or its publication. This work was supported by the New Faculty Research Support Grant from Gyeongsang National University in 2025, GNU-NFRSG-0021. This work was also supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT), RS-2025-23323982.

## References

Abascal, F., Harvey, L. M. R., Mitchell, E., Lawson, A. R. J., Lensing, S. V., Ellis, P., et al. (2021). Somatic mutation landscapes at single-molecule resolution. *Nature* 593 (7859), 405–410. doi:10.1038/s41586-021-03477-4

Adameyko, I., and Lallemend, F. (2010). Glial versus melanocyte cell fate choice: Schwann cell precursors as a cellular origin of melanocytes. *Cell Mol. Life Sci.* 67 (18), 3037–3055. doi:10.1007/s00018-010-0390-y

Almeida, M., and Ranisch, R. (2022). Beyond safety: mapping the ethical debate on heritable genome editing interventions. *Humanit. Soc. Sci. Commun.* 9 (1), 139. doi:10.1057/s41599-022-01147-y

Bae, T., Tomasini, L., Mariani, J., Zhou, B., Roychowdhury, T., Franjic, D., et al. (2018). Different mutational rates and mechanisms in human cells at pregastrulation and neurogenesis. *Science* 359 (6375), 550–555. doi:10.1126/science.aan8690

Behjati, S., Huch, M., van Boxtel, R., Karthaus, W., Wedge, D. C., Tamuri, A. U., et al. (2014). Genome sequencing of normal cells reveals developmental lineages and mutational processes. *Nature* 513 (7518), 422–425. doi:10.1038/nature13448

Bizzotto, S., Dou, Y., Ganz, J., Doan, R. N., Kwon, M., Bohrson, C. L., et al. (2021). Landmarks of human embryonic development inscribed in somatic mutations. *Science* 371 (6535), 1249–1253. doi:10.1126/science.abe1544

Blankpain, C., and Simons, B. D. (2013). Unravelling stem cell dynamics by lineage tracing. *Nat. Rev. Mol. Cell Biol.* 14 (8), 489–502. doi:10.1038/nrm3625

Boysen, G., Alexandrov, L. B., Rahbari, R., Nookaew, I., Ussery, D., Chao, M.-R., et al. (2025). Investigating the origins of the mutational signatures in cancer. *Nucleic Acids Res.* 53 (1), gkae1303. doi:10.1093/nar/gkae1303

Breuss, M. W., Yang, X., Schlachetzki, J. C. M., Antaki, D., Lana, A. J., Xu, X., et al. (2022). Somatic mosaicism reveals clonal distributions of neocortical development. *Nature* 604 (7907), 689–696. doi:10.1038/s41586-022-04602-7

Bukowy-Bierylo, Z. (2021). Long-term differentiating primary human airway epithelial cell cultures: how far are we? *Cell Commun. Signal.* 19 (1), 63. doi:10.1186/s12964-021-00740-z

Cagan, A., Baez-Ortega, A., Brzozowska, N., Abascal, F., Coorens, T. H. H., Sanders, M. A., et al. (2022). Somatic mutation rates scale with lifespan across mammals. *Nature* 604 (7906), 517–524. doi:10.1038/s41586-022-04618-z

Choi, S. H., Ku, E. J., Choi, Y. A., and Oh, J. W. (2023). Grave-to-cradle: human embryonic lineage tracing from the postmortem body. *Exp. and Mol. Med.* 55 (1), 13–21. doi:10.1038/s12276-022-00912-y

Coorens, T. H. H., Moore, L., Robinson, P. S., Sanghvi, R., Christopher, J., Hewinson, J., et al. (2021a). Extensive phylogenies of human development inferred from somatic mutations. *Nature* 597 (7876), 387–392. doi:10.1038/s41586-021-03790-y

Coorens, T. H. H., Oliver, T. R. W., Sanghvi, R., Sovio, U., Cook, E., Vento-Tormo, R., et al. (2021b). Inherent mosaicism and extensive mutation of human placentas. *Nature* 592 (7852), 80–85. doi:10.1038/s41586-021-03345-1

Coorens, T. H. H., Spencer Chapman, M., Williams, N., Martincorena, I., Stratton, M. R., Nangalia, J., et al. (2024). Reconstructing phylogenetic trees from genome-wide somatic mutations in clonal samples. *Nat. Protoc.* 19 (6), 1866–1886. doi:10.1038/s41596-024-00962-8

Coorens, T. H. H., Collard, G., Jung, H., Wang, Y., Moore, L., Hooks, Y., et al. (2025). The somatic mutation landscape of normal gastric epithelium. *Nature* 640 (8058), 418–426. doi:10.1038/s41586-025-08708-6

De, S. (2011). Somatic mosaicism in healthy human tissues. *Trends Genet.* 27 (6), 217–223. doi:10.1016/j.tig.2011.03.002

Dou, Y., Gold, H. D., Luquette, L. J., and Park, P. J. (2018). Detecting somatic mutations in normal cells. *Trends Genet.* 34 (7), 545–557. doi:10.1016/j.tig.2018.04.003

Ellis, P., Moore, L., Sanders, M. A., Butler, T. M., Brunner, S. F., Lee-Six, H., et al. (2021). Reliable detection of somatic mutations in solid tissues by laser-capture microdissection and low-input DNA sequencing. *Nat. Protoc.* 16 (2), 841–871. doi:10.1038/s41596-020-00437-6

Fasching, L., Jang, Y., Tomasi, S., Schreiner, J., Tomasini, L., Brady, M. V., et al. (2021). Early developmental asymmetries in cell lineage trees in living individuals. *Science* 371 (6535), 1245–1248. doi:10.1126/science.abe0981

Freed, D., Stevens, E. L., and Pevsner, J. (2014). Somatic mosaicism in the human genome. *Genes (Basel)* 5 (4), 1064–1094. doi:10.3390/genes5041064

Gonzalez-Pena, V., Natarajan, S., Xia, Y., Klein, D., Carter, R., Pang, Y., et al. (2021). Accurate genomic variant detection in single cells with primary template-directed amplification. *Proc. Natl. Acad. Sci.* 118 (24), e2024176118. doi:10.1073/pnas.2024176118

Horak, P., Griffith, M., Danos, A. M., Pitel, B. A., Madhavan, S., Liu, X., et al. (2022). Standards for the classification of pathogenicity of somatic variants in cancer (oncogenicity): joint recommendations of clinical genome resource (ClinGen), cancer genomics consortium (CGC), and variant interpretation for cancer consortium (VICC). *Genet. Med.* 24 (5), 986–998. doi:10.1016/j.gim.2022.01.001

Huang, A. Y., and Lee, E. A. (2021). Identification of somatic mutations from bulk and single-cell sequencing data. *Front. Aging* 2, 800380. doi:10.3389/fragi.2021.800380

## Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Huang, L., Ma, F., Chapman, A., Lu, S., and Xie, X. S. (2015). Single-cell whole-genome amplification and sequencing: methodology and applications. *Annu. Rev. Genomics Hum. Genet.* 16, 79–102. doi:10.1146/annurev-genom-090413-025352

Johnson, M. L., Sima, C. S., Chafft, J., Paik, P. K., Pao, W., Kris, M. G., et al. (2013). Association of KRAS and EGFR mutations with survival in patients with advanced lung adenocarcinomas. *Cancer* 119 (2), 356–362. doi:10.1002/cncr.27730

Ju, Y. S., Martincorena, I., Gerstung, M., Petljak, M., Alexandrov, L. B., Rahbari, R., et al. (2017). Somatic mutations reveal asymmetric cellular dynamics in the early human embryo. *Nature* 543 (7647), 714–718. doi:10.1038/nature21703

Junyent, S., Meglicki, M., Vetter, R., Mandelbaum, R., King, C., Patel, E. M., et al. (2024). The first two blastomeres contribute unequally to the human embryo. *Cell* 187 (11), 2838–2854.e2817. doi:10.1016/j.cell.2024.04.029

Kalef-Ezra, E., Turan, Z. G., Perez-Rodriguez, D., Bommann, I., Behera, S., Morley, C., et al. (2024). Single-cell somatic copy number variants in brain using different amplification methods and reference genomes. *Commun. Biol.* 7 (1), 1288. doi:10.1038/s42003-024-06940-w

Kapadia, C. D., Williams, N., Dawson, K. J., Watson, C., Yousefzadeh, M. J., Le, D., et al. (2025). Clonal dynamics and somatic evolution of haematopoiesis in mouse. *Nature* 641 (8063), 681–689. doi:10.1038/s41586-025-08625-8

Kennedy, S. R., Schmitt, M. W., Fox, E. J., Kohrn, B. F., Salk, J. J., Ahn, E. H., et al. (2014). Detecting ultralow-frequency mutations by duplex sequencing. *Nat. Protoc.* 9 (11), 2586–2606. doi:10.1038/nprot.2014.170

Kim, J. Y., Lee, E., Park, K., Park, W. Y., Jung, H. H., Ahn, J. S., et al. (2017). Clinical implications of genomic profiles in metastatic breast cancer with a focus on TP53 and PIK3CA, the most frequently mutated genes. *Oncotarget* 8 (17), 27997–28007. doi:10.18632/oncotarget.15881

Kim, J. H., Hwang, S., Son, H., Kim, D., Kim, I. B., Kim, M.-H., et al. (2022). Analysis of low-level somatic mosaicism reveals stage and tissue-specific mutational features in human development. *PLOS Genet.* 18 (9), e1010404. doi:10.1371/journal.pgen.1010404

Kwon, S. G., Bae, G. H., Hong, J. H., Choi, J. W., Choi, J. H., Lim, N. S., et al. (2024). Comprehensive analysis of somatic mutations and structural variations in domestic pig. *Mamm. Genome* 35 (4), 645–656. doi:10.1007/s00335-024-10058-z

Lee-Six, H., Øbro, N. F., Shepherd, M. S., Grossmann, S., Dawson, K., Belmonte, M., et al. (2018). Population dynamics of normal human blood inferred from somatic mutations. *Nature* 561 (7724), 473–478. doi:10.1038/s41586-018-0497-0

Lodato, M. A., and Vijg, J. (2022). Editorial: somatic mutations, genome mosaicism and aging. *Front. Aging* 3, 1115408. doi:10.3389/fragi.2022.1115408

Lodato, M. A., Woodworth, M. B., Lee, S., Evrony, G. D., Mehta, B. K., Karger, A., et al. (2015). Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* 350 (6256), 94–98. doi:10.1126/science.aab1785

Loeb, L. A., and Cheng, K. C. (1990). Errors in DNA synthesis: a source of spontaneous mutations. *Mutat. Res.* 238 (3), 297–304. doi:10.1016/0165-1110(90)90021-3

Matas, J., Kohrn, B., Fredrickson, J., Carter, K., Yu, M., Wang, T., et al. (2022). Colorectal cancer is associated with the presence of cancer driver mutations in normal Colon. *Cancer Res.* 82 (8), 1492–1502. doi:10.1158/0008-5472.Can-21-3607

Menon, V., and Brash, D. E. (2023). Next-generation sequencing methodologies to detect low-frequency mutations. *Catch Me If You Can Mutat. Res. - Rev. Mutat. Res.* 792, 108471. doi:10.1016/j.mrrev.2023.108471

Mitchell, E., Spencer Chapman, M., Williams, N., Dawson, K. J., Mende, N., Calderbank, E. F., et al. (2022). Clonal dynamics of haematopoiesis across the human lifespan. *Nature* 606 (7913), 343–350. doi:10.1038/s41586-022-04786-y

Moeller, M. E., Père, N. V. M., Werner, B., and Huang, W. (2023). Measures of genetic diversification in somatic tissues at bulk and single cell resolution. *eLife Sci. Publ. Ltd.* doi:10.7554/eLife.89780

Mohiuddin, M., Kooy, R. F., and Pearson, C. E. (2022). *De novo* mutations, genetic mosaicism and human disease. *Front. Genet.* 13, 13–2022. doi:10.3389/fgene.2022.983668

Moore, L., Cagan, A., Coorens, T. H. H., Neville, M. D. C., Sanghvi, R., Sanders, M. A., et al. (2021). The mutational landscape of human somatic and germline cells. *Nature* 597 (7876), 381–386. doi:10.1038/s41586-021-03822-7

Osorio, F. G., Rosendahl Huber, A., Oka, R., Verheul, M., Patel, S. H., Hasaart, K., et al. (2018). Somatic mutations reveal lineage relationships and age-related mutagenesis in human hematopoiesis. *Cell Rep.* 25 (9), 2308–2316.e2304. doi:10.1016/j.celrep.2018.11.014

Park, S., Mali, N. M., Kim, R., Choi, J.-W., Lee, J., Lim, J., et al. (2021). Clonal dynamics in early human embryogenesis inferred from somatic mutation. *Nature* 597 (7876), 393–397. doi:10.1038/s41586-021-03786-8

Pennington, K. P., Walsh, T., Harrell, M. I., Lee, M. K., Pennil, C. C., Rendi, M. H., et al. (2014). Germline and somatic mutations in homologous recombination genes predict platinum response and survival in ovarian, fallopian tube, and peritoneal carcinomas. *Clin. Cancer Res.* 20 (3), 764–775. doi:10.1158/1078-0432.Ccr-13-2287

Pleasance, E. D., Cheetham, R. K., Stephens, P. J., McBride, D. J., Humphray, S. J., Greenman, C. D., et al. (2010). A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* 463 (7278), 191–196. doi:10.1038/nature08658

Rahal, Z., Scheet, P., and Kadara, H. (2024). Somatic mutations in normal tissues: calm before the storm. *Cancer Discov.* 14 (4), 605–609. doi:10.1158/2159-8290.Cd-23-1508

Rajendran, B. K., and Deng, C. X. (2017). Characterization of potential driver mutations involved in human breast cancer by computational approaches. *Oncotarget* 8 (30), 50252–50272. doi:10.18632/oncotarget.17225

Rawlins, E. L., Okubo, T., Xue, Y., Brass, D. M., Auten, R. L., Hasegawa, H., et al. (2009). The role of Scgb1a1+ clara cells in the long-term maintenance and repair of lung airway, but not alveolar, epithelium. *Cell Stem Cell* 4 (6), 525–534. doi:10.1016/j.stem.2009.04.002

Ren, P., Dong, X., and Vijg, J. (2022). Age-related somatic mutation burden in human tissues. *Front. Aging Volume*, 3–2022. doi:10.3389/fragi.2022.1018119

Rzepecka, I. K., Szafron, L. M., Stys, A., Felisiak-Golabek, A., Podgorska, A., Timorek, A., et al. (2017). Prognosis of patients with BRCA1-associated ovarian carcinomas depends on TP53 accumulation status in tumor cells. *Gynecol. Oncol.* 144 (2), 369–376. doi:10.1016/j.ygyno.2016.11.028

Simeonov, K. P., Byrns, C. N., Clark, M. L., Norgard, R. J., Martin, B., Stanger, B. Z., et al. (2021). Single-cell lineage tracing of metastatic cancer reveals selection of hybrid EMT states. *Cancer Cell* 39 (8), 1150–1162.e1159. doi:10.1016/j.ccr.2021.05.005

Simons, B. D., and Clevers, H. (2011). Strategies for homeostatic stem cell self-renewal in adult tissues. *Cell* 145 (6), 851–862. doi:10.1016/j.cell.2011.05.033

Snippert, H. J., van der Flier, L. G., Sato, T., van Es, J. H., van den Born, M., Kroon-Veenboer, C., et al. (2010). Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell* 143 (1), 134–144. doi:10.1016/j.cell.2010.09.016

Spencer Chapman, M., Ranzoni, A. M., Myers, B., Williams, N., Coorens, T. H. H., Mitchell, E., et al. (2021). Lineage tracing of human development through somatic mutations. *Nature* 595 (7865), 85–90. doi:10.1038/s41586-021-03548-6

Spencer Chapman, M., Wilk, C. M., Boettcher, S., Mitchell, E., Dawson, K., Williams, N., et al. (2024). Clonal dynamics after allogeneic haematopoietic cell transplantation. *Nature* 635 (8040), 926–934. doi:10.1038/s41586-024-08128-y

Sulston, J. E., Schierenberg, E., White, J. G., and Thomson, J. N. (1983). The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol.* 100 (1), 64–119. doi:10.1016/0012-1606(83)90201-4

Tang, Y., and Fan, Y. (2024). Combined KRAS and TP53 mutation in patients with colorectal cancer enhance chemoresistance to promote postoperative recurrence and metastasis. *BMC Cancer* 24 (1), 1155. doi:10.1186/s12885-024-12776-8

Uchimura, A., Matsumoto, H., Satoh, Y., Minakuchi, Y., Wakayama, S., Wakayama, T., et al. (2022). Early embryonic mutations reveal dynamics of somatic and germ cell lineages in mice. *Genome Res.* 32 (5), 945–955. doi:10.1101/gr.276363.121

Vogt, W. (1929). Gestaltungsanalyse am Amphibienkeim mit Örtlicher Vitalfärbung: II. Teil. Gastrulation und Mesodermbildung bei Urodelen und Anuren. *Wilhelm Roux Arch. Entwickl Mech. Org.* 120 (1), 384–706. doi:10.1007/bf02109667

Wang, T., Chen, X., Wang, K., Ju, J., Yu, X., Wang, S., et al. (2023). Cre-loxP-mediated genetic lineage tracing: unraveling cell fate and origin in the developing heart. *Front. Cardiovasc Med.* 10, 1085629. doi:10.3389/fcvm.2023.1085629

Wang, J., Li, H., and Liu, H. (2024). Prioritizing clinically significant lung cancer somatic mutations for targeted therapy through efficient NGS data filtering system. *AMIA Jt. Summits Transl. Sci. Proc.* 2024, 305–313.

Werner, B., and Sotteriva, A. (2018). Variation of mutational burden in healthy human tissues suggests non-random strand segregation and allows measuring somatic mutation rates. *PLOS Comput. Biol.* 14 (6), e1006233. doi:10.1371/journal.pcbi.1006233

Whitman, C. O. (1878). The embryology of Clepsine1. *J. Cell Sci.* s2-18 (71), 215–315. doi:10.1242/jcs.18.71.2125

Woo, K., and Fraser, S. E. (1995). Order and coherence in the fate map of the zebrafish nervous system. *Development* 121 (8), 2595–2609. doi:10.1242/dev.121.8.2595

Yaacov, A., Rosenberg, S., and Simon, I. (2023). Mutational signatures association with replication timing in normal cells reveals similarities and differences with matched cancer tissues. *Sci. Rep.* 13 (1), 7833. doi:10.1038/s41598-023-34631-9

Youk, J., Kwon, H. W., Kim, R., and Ju, Y. S. (2021). Dissecting single-cell genomes through the clonal organoid technique. *Exp. Mol. Med.* 53 (10), 1503–1511. doi:10.1038/s12276-021-00680-1

Zhou, B., Ma, Q., Rajagopal, S., Wu, S. M., Domian, I., Rivera-Feliciano, J., et al. (2008). Epicardial progenitors contribute to the cardiomyocyte lineage in the developing heart. *Nature* 454 (7200), 109–113. doi:10.1038/nature07060