DHF 1: Needs Assessment

1.1 Client & Project Introduction

The Yachie Laboratory is one of the foremost synthetic biology laboratories, with a primary focus on understanding cell development for translation into regenerative medicine and therapeutic applications. Our client, Dr. Nozomi Yachie and his lab, have successfully developed "DNA tapes" to explain various aspects of complex cellular systems, including cell lineage tracing. Most recently, Dr. Yachie has become invested in retrospectively quantifying cell division and gene expression in mammalian cells. Based on the same principles as before, high throughput sequencing technologies will be paired with the four pillars of DNA recording to understand the history of a cell on a molecular level. These pillars are 1) memory, to record and keep track of DNA bases, 2) sensor, to detect molecular and biological events, 3) writer, to enable genome editing, and lastly 4) reader, to read the data in an interpretable and accessible format. A need for designing such genetic circuits exists because most synthetic circuit technologies for DNA event recordings are static, and dynamic systems have yet to be thoroughly developed. The scalability of such a technology also poses a problem that is yet to be addressed. Such synthetic circuits have immense potential to improve health care, diagnostics, regenerative medicine and education. The goal of the Yachie lab is to map the entire cell lineage and differentiation patterns during development in a mouse.

Our first meeting with Dr.Yachie, and members of his lab, took place on September 29th, 2021, where we defined the needs and scope of the project. A summary of our first meeting is provided in Appendix A of DHF 13 Section 13.2.4. The end goal for this project is to build a scalable synthetic gene circuit with the ability to record and count the number of times each cell in a heterogeneous mammalian cell population has undergone a cell division. Additionally, Dr. Yachie would like us to retrospectively approximate the expression profile of a target gene in each cell. The latter goal would be more time-consuming and may not be feasible in the time scale provided, so it will be undertaken only if the first goal is successful.

1.2 Stakeholder Investigation

Table 1 identifies key stakeholders in the development of our final system. We plan to keep a consistent line of contact with one stakeholder in each stakeholder group (i.e., clients, end-users, rightsholders, and others), and to consult our other stakeholders when necessary. As a first step, we will identify potential contacts and address our most critical questions.

Table 1: Key stakeholders, their needs and values, and an investigation plan with critical questions.

Stakeholder	Needs & Values	Investigation Plan & Critical Questions
Clients		
Dr. Nozomu Yachie	• Contribute to the scientific community	Plan: Continue to consult on project outcomes and goals when needed.

	● Use synthetic and computational biology to advance regenerative medicine ● Devise an approach to record quantitative biological events	Questions: ● What is the current state of the project? ● What are your goals and future directions for the project? ● How much supervision and training will be provided? ● Who will be our direct point of contact? ● How regularly will we debrief? ● What resources will we have access to?
Instructional team	● Assist in the completion of capstone design projects	Plan: Continue to attend weekly meetings and keep up-to-date meeting minutes. Questions: ● What matters should be discussed first with instructors and teaching assistants? ● What matters can be addressed directly with the client?
Granting agencies	● Complete the project within a stipulated timeframe and budget	Plan: Research more about potential granting agencies for the long-term completion of the project and what their requirements are. Questions: ● How much funding is available? ● What are the requirements to secure the funding? ● What limitations does this impose on the ability to self-regulate?
End-users		
Other members and collaborators[1] of the Yachie laboratory	● Develop and apply genome editing technologies and bioinformatics tools	Plan: Establish a line of contact for potential collaboration on the project beginning with our existing contacts (the Zandstra Lab, Kieffer Lab, etc.). Questions: ● What is their background and research focus? ● How can this tool assist them in their research?

[1] Aburatani, Landry, Nishida, Nureki, Ota, and Sasaki laboratories.

		● What is lacking in existing technologies?
Other research institutions	● Pioneer scientific advances ● Validate or expand upon work from other laboratories	Questions: ● Is there a competition to develop this type of technology?
Academic publishers	● Disseminate knowledge to members of the scientific community	Plan: Research publications related to cell division recording and how this information was disseminated. Questions: ● What is the intended audience? ● Is this a method paper?
Rightsholders		
Software developers	● Receive recognition for their work and intellectual property	Questions: ● Will we be integrating other tools into this technology? ● How do we credit contributors?
Animal rights activists	● Protect the rights of animals used in research studies	Plan: Consult with the Yachie Lab on what long-term *in vivo* work might be done for the project. Questions: ● Does this project require *in vivo* validation with mouse models?
Others		
Suppliers	● Provide the tools and materials to conduct research	Plan: Research different supplier streams for synthetic gene circuits and consult with the Yachie Lab about available resources. Questions: ● What facilities will we have access to? ● Who will keep track of inventory and purchase new orders?
Policymakers	● Support the advancement of scientific research	● Is the government currently involved in funding?

	• Safeguard public welfare	• Is there a goal to prove the safety or efficacy of these technologies as a means of increasing support from policymakers?
Healthcare providers	• Improve current methods for diagnosing and treating genetic diseases	Plan: Research how a cell division recording system would benefit healthcare providers and what limitations there may be in the implementation of the system in their working environment. Questions: • Will this product be directly utilized by healthcare providers? • How can healthcare providers make better informed decisions using these tools? • When and how does knowledge translation take place?

1.3 Value Proposition

The utilization of a cell-embedded genetic circuit for recording cell divisions can help us capture the cellular lineage and differentiation histories of a heterogeneous population of cells at high-resolution. This system can help revolutionize our understanding of cell fate determination, development, and the disorders that may arise for erroneous gene regulatory events. Through the usage of high-capacity "DNA tapes", molecular sensors of cell lineage, information "writers" utilizing CRISPR genome editing technology in combination with sequencing and computing technologies, scientists in industry and academia can utilize the high-resolution, whole-body dataset visualizing cell lineage and differentiation of entire mammalian (mouse) development.

1.4 Identifying Alternative Solutions

Table 2: Existing solutions to cell division recording.

Solution	Strengths	Weaknesses	Source
Evolving Genetic Barcodes	• Enables a versatile platform for *in vivo* barcoding and lineage tracing in a mammal. • The barcode elements are genome integrated, stable, and heritable that can be activated by simply crossing with another cell line.	• The lineage of the cells was derived based on the similarity of the barcode sequence, which is not completely accurate. • It requires killing lots of cells and sequencing massive numbers of barcodes to predict the lineage, which is time consuming.	[Link]

	• hgRNAs are scattered throughout the genome, which circumvents large deletion events. • The scattered hgRNA accumulate mutations independently which creates a large diversity of barcodes. • This remarkable diversity is adequate for uniquely barcoding every cell.	• It requires read-out and machine learning technologies with high-throughput capacity. • Genetic manipulation of an individual germ cell is challenging which includes genome localization, insertion, and transplantation.	
intMEMOIR	• Editing can operate in different organisms and contexts. • It simultaneously analyzes live single cell lineage, spatial organization, and gene expression data in the same tissue. • The system can be applied to reconstruct the dynamic activity histories of signaling pathways and transcription factors by introducing more distinct integrases.	• The number of combinations is insufficient to represent all individual cells uniquely. • The fluorescent marker reveals the relation of a pile of cells rather than each single cell. • There are limited types of fluorescence with limited color spectrum, therefore it cannot generate all kinds of colors to represent all cells.	[Link]
Sparse Retroviral Labeling	• Retroviral library labeling is an advantageous method for determining lineage relationships both *in vivo* and *ex vivo*. • Viral infection can precisely target specific cells that express the viral receptor. • The cells can be directly visualized by its fluorescent emission rather than sequencing. • Its gene editing procedures are less complex.	• Only cells with the capacity to divide will propagate the barcode to progeny. • Retroviral vectors can spontaneously silence (no histochemical features expressed). • Barcode tag recovery from single cells can be challenging. • *Ex vivo* progenitor divisions and behavior may differ from that observed *in vivo*. • It does not mark each cell uniquely.	[Link]

	• It helps to monitor and record cell division.	• It does not record the number of divisions automatically. • It does not produce detailed cell lineage for each individual cell.	
Plasmid Transfection Labeling	• Plasmid Transfection Labeling is an advantageous method for determining lineage relationships both *in vivo* and *in vitro*. • The cells can be directly visualized by its fluorescent emission rather than sequencing. • Its gene editing procedures are less complex. • It helps to monitor and record cell division.	• Plasmid DNA is not integrated into the genome of the progenitor which becomes diluted or inactivated in progeny after serial cellular divisions. Plasmid electroporation techniques, therefore, are transient and fail to label the entire lineage. • It does not record the number of divisions automatically. • It does not mark each cell uniquely. • It does not produce detailed cell lineage for each individual cell.	[Link]
Genetic Recombin -ation	• Genetic Recombination can be used to detect the expression of a particular gene or the occurrence of a particular event. • The cells can be directly visualized by its fluorescent emission rather than sequencing. • Its gene editing procedures are less complex. • It helps to monitor and record cell division.	• Inducible Cre systems are leaky. • It does not record the number of divisions automatically. • It does not produce detailed cell lineage for each individual cell. • It does not mark each cell uniquely. • It requires *in vivo* experiments to combine recombinase and target genes together.	[Link]
Somatic Mutations	• No additional gene manipulation techniques are required, the sample can be directly used for sequencing. • It helps to monitor and record cell division. • Somatic mutation can not only count for the	• It does not record the number of divisions automatically. • It does not produce detailed cell lineage for each individual cell. • Somatic mutation happens stochastically at a low rate when a cell divides, therefore, it does not target a specific cell	[Link]

	differences among individual cells but also estimate their affinity.	type or mark an individual cell uniquely.	
The Bacterial Nano-recorder	• The structure and mechanisms of bacteria are well studied, it is easier to engineer its constitution and alter its function. • The detection of activated cells becomes much easier compared to previous bacterial recorders. • It can perform timestamping since the length of the cell is proportional to the time following exposure. • It can possibly record the expression of a particular gene, but it needs to be modified to record the amount of expression.	• The system is self-limiting: cells stop generating viable progeny after exposure to the chemical stimulus, which impacts the cell behavior significantly. • It might be hard to record the cell division since it requires a specific chemical response. • It is not heritable but transmissible, recording cell lineage is less possible, but it can record cell behavior for a single cell. • The bacteria may mutate into an infectious state to cause disease.	[Link]
Microphoto-graphic Recording	• It does not involve gene manipulation. Therefore, it is much more convenient and intuitive to use. • It can approximate the division time for each cell. • Cell line family trees can be traced in the recording. • This technique is suitable for a lab setting without *in vivo* tests. • It does not need to kill the cell to extract information.	• This technique is not applicable to record cell lineage *in vivo*. The experiment largely depends on cell culture, and cell behaviors may be affected. • It can only record the lineage for a limited number of cells. • It does not count the cell division automatically. • It does not record the approximate amount of gene expression of a target gene.	[Link]

The evolving genetic barcodes are created using a specific DNA sequence that encodes an engineered RNA molecule called a homing guide RNA (hgRNA). With the presence of enzyme Cas9, the hgRNA will guide the Cas9 to its original hgRNA coded genome and cut the genome. The DNA repair mechanism will introduce genetic mutations in the hgRNA sequence when repairing the cut. These mutations accumulate over time to build a unique barcode with division

history recorded for each cell. Eventually, the lineage between different cells was determined by comparing the level of mutations they have.

The full name of intMEMOIR is integrase-editable memory by engineered mutagenesis with optical in situ readout, which is a digital, image-readable lineage recording system based on site-specific serine integrases, such as Bxb1. This design is based on a barcode that contains an array of 10 independent three-state genetic memory elements that can be integrated at defined genomic sites for germline heritability. Each element can be digitally, stochastically, and irreversibly edited, allowing up to 59,049 distinct combinations. These digital states can be read out alongside endogenous transcripts using fluorescence in situ hybridization (FISH) methods.

Sparse Retroviral Labeling uses retroviral vector-mediated gene transfer that allows viruses to introduce recombinant DNA into the genome of a host cell. Viruses are applied at limiting dilutions to label single founder cells. The integrated exogenous DNA encodes a histochemical or fluorescent protein that can be easily detected, which will be inherited by all the descendants of the infected cell. Therefore, the progeny of virally infected cells are labeled, which elucidates cell fate choices within that clone. Histological and morphological analyses of these cells allow for post hoc fate mapping within a clonally related cell population using time-lapse imaging. Individual progenitors that have been labeled with fluorescent reporter genes are visualized using confocal microscopy for multiple cellular divisions.

Plasmid Transfection Labeling uses DNA plasmid transfection to insert reporter transgenes for cell labeling and fate mapping. A reporter transgene, such as GFP, is then carried by the progenitor cell and passed on to subsequent daughter cells. Lipofection is a common lipid-based system used to transfect the cells for *in vivo* and *in vitro* lineage. Electroporation is an alternative non-viral delivery method that uses electrical fields to increase cell membrane permeability to recombinant DNA. In addition, *in utero* electroporation (IUE) has proved to be an efficient technique to introduce recombinant DNA plasmids into neural progenitors *in vivo*.

Genetic recombination activates the expression of a conditional reporter gene based on the expression of recombinase enzymes. There are two genetically encoded, site-specific recombinase recombination systems: Cre-loxP and FLP-FRT. In the Cre-loxP system, the organism is engineered to express Cre recombinase under the control of a chosen promoter, the expression of Cre was limited to a specific tissue or cell type. These lines are then crossed with a second line in which a reporter transgene, such as lacZ or GFP, is preceded by a loxP-flanked transcriptional stop (loxP-STOP-loxP) cassette. The STOP sequence will be excised to express the reporter transgene when Cre recombinase is present. Genetic recombination can be temporally controlled by using an inducible Cre system. In an inducible system, Cre recombinase is fused to a human hormone receptor and activated only in the presence of an anti-hormone.

Somatic mutations, such as copy-number variation and single-nucleotide polymorphism, are discovered by research as a spontaneous cell lineage tracing media. There are endogenous retroelements that constitute much of the human genome, which principally consist of long

interspersed nuclear element 1 (L1; also known as LINE-1) elements. A very small number of these L1 elements retain the ability to mobilize in humans and can insert into a new genomic location during somatic cell division, which may potentially contribute to somatic diversity. The frequency of L1 retrotransposition events is low, but it can still illustrate the spatial distribution patterns of clonal cell lineage. Similar retro transposition events can be found using a digital droplet PCR assay.

The Bacterial Nanorecorder was designed to determine both the type and the time at which a brief chemical exposure event has occurred. A genetic memory element was used such as the bistable toggle switch, a synthetic genetic circuit that generates a sustained cellular response to a transient stimulus. It responds to a brief event or stimulus, maintains a memory of that event/stimulus, and time stamps that event or stimulus. In other words, a synthetic bacterium was created to register a long-term phenotypic response following short-term exposure to a chemical and utilize the elicited phenotypic response to estimate the time elapsed after chemical exposure, such as a uniform cell elongation, with cell length being directly proportional to the time elapsed.

Microphotographic Recording used two types of microscopes: stereomicroscopes and an inverted Reichert microscope. Stereomicroscopes were used for selecting the cells whose growth was followed in the experiments. The location of the selected cells was marked with a felt-tip marker. Cell multiplication was photographically recorded with the aid of either a stereomicroscope or the inverted microscope. Division time was estimated from the time between two recordings. For a time-lapse photomicroscope, the fully automatic Robot Star11 camera was attached to the microscope. For a timelapse video recording microscopy, an Ikegami television camera was attached to the Nikon stereomicroscope and connected to a video recorder.

1.5 Use Cases

Figure 1 shows the paths for multiple use cases of the system split into three phases: initialization, data gathering, and data distribution. The phases will likely be iterative as research groups optimize the system for their applications. The main actor of the system is a researcher. A researcher performs tasks of the system in all three phases, which are all associated with the development team that maintains and distributes the system.

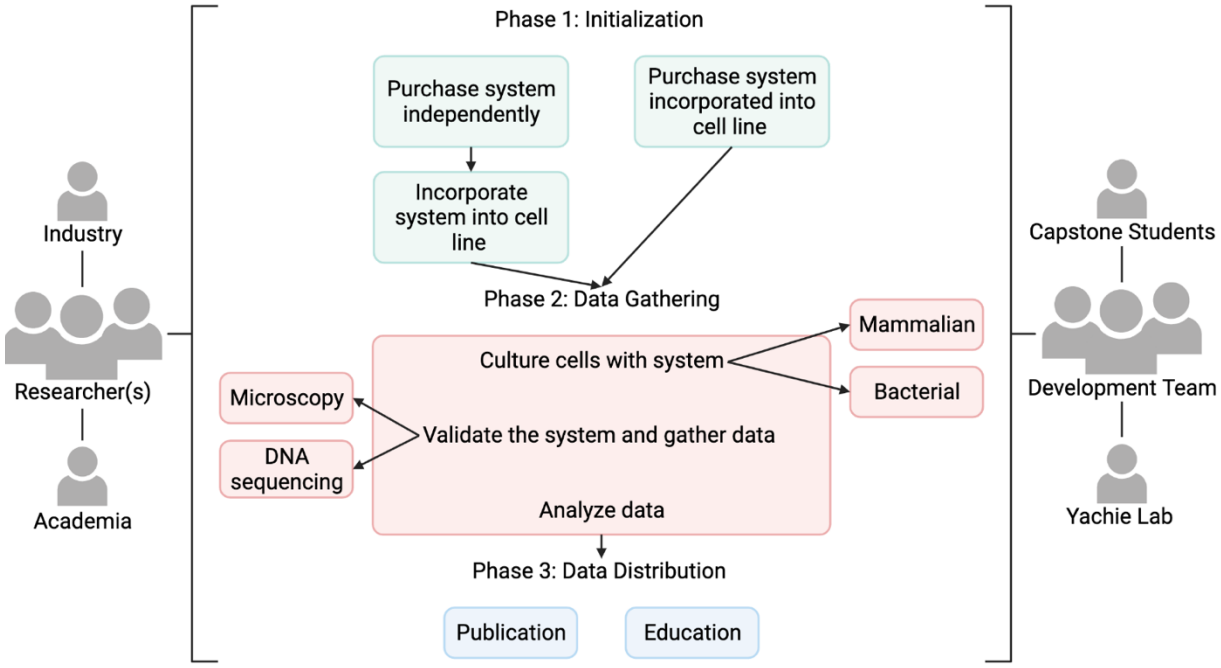

Figure 1: Use-case diagram.

The data distribution phase is the main purpose of the system – informing the scientific community about how the data can be applied. For example, researchers can use a dynamic cell division counter in quality control of stem cells to see how many cell divisions occur before a sample's mutational burden affects its cell fate and function. Werner *et al.* inferred the mutation and survival rates per cell division of healthy brain development and hematopoiesis using single time point sequencing data (1). Given a cell division count, and the rates found by Werner *et al.*, researchers can identify the number of mutations and survival probability of a cell sample. Cell division counts can also be used to find the cell division rate of a sample during specific culture conditions. Fluctuations in cell division rates are accompanied by changes in gene expression, which can affect biomolecule concentrations. Bertaux *et al.* found a "'U' shaped dependency of [gene expression] noise on division rate" (2). Using both cell division rate and cumulative gene expression, researchers can then identify how expression of a gene varies as a function of cell division rate. In addition, cell division counts can provide insights into developmental biology, such as the use of iCOUNT by Denoth-Lippuner *et al. t*o characterize the division patterns of radial glia cells in the cortex of mice (3).

References

1. Werner B, Case J, Williams MJ, Chkhaidze K, Temko D, Fernández-Mateos J, et al. Measuring single cell divisions in human tissues from multi-region sequencing data. Nat Commun. 2020;11(1):1–9.
2. Bertaux F, Marguerat S, Shahrezaei V. Division rate, cell size and proteome allocation: impact on gene expression noise and implications for the dynamics of genetic circuits. R Soc Open Sci. 2018;5(3):172234.
3. Denoth-Lippuner A, Jaeger BN, Liang T, Royall LN, Chie SE, Buthey K, et al. Visualization of individual cell division history in complex tissues using iCOUNT. Cell Stem Cell. 2021;28:1–15.