Pioneering work in the field of genome engineering, particularly by developing new methods to record genome expressions in living cells

Randall J Platt

ETH Zürich

Introduction

What are the main differences between a photograph and a video? Photographs record a single point in time and videos continuously record a sequence of events. While the content and interpretation of a photograph is heavily reliant on a single moment in time, a video is not. The tools that scientists have to understand the molecular and cellular world around them are most like film cameras – producing single snapshots to describe dynamic processes. Towards the goal of continuously recording molecular events within cells, my laboratory recently developed 'transcriptional recording', an approach that employs CRISPR spacer acquisition from RNA to capture and convert intracellular RNAs into DNA, permanently storing transcriptional information in the DNA of living cells. The newly acquired sequences serve as transcriptional records, which are retrievable via deep sequencing and can be leverage to reconstruct cellular histories. This technology provides an entirely new mode of interrogating dynamic biological and physiological processes and opens up numerous avenues for future work.

The question: How do we preserve dynamic and transient biological information throughout time?

A fundamental challenge in biology is to understand how cells function and integrate complex molecular information to perform different behaviors. This challenge has motivated the creation of numerous technologies facilitating detailed intracellular observations at the level of DNA, RNA, protein, and metabolites[1]. In particular, RNA sequencing (RNA-seq) facilitates transcriptome quantification within multiple or single cells, revealing the molecular signatures of cell behaviors, states, and types with unprecedented detail[2, 3]. Despite the power of these approaches, they require destructive methods and therefore observations are limited to a few snapshots in time or select asynchronous cellular processes. One provocative solution to this is to introduce DNA writing and molecular recording devices within cells that enable the encoding, storage, and retrieval of molecular information[4-6].

The microbial adaptive immune system CRISPR–Cas embodies the ideal DNA writing and molecular recording system. CRISPR loci are comprised of CRISPR-associated (*cas*) genes and CRISPR arrays, which store molecular memories of prior infections in the form of short nucleic acid segments (spacers) acquired from foreign genetic elements separated by direct repeats (DRs)[7, 8]. *In vitro* and *in vivo* experiments using the *E. coli* Type I-E CRISPR–Cas system, support a mechanism where Cas1 and Cas2 are the only Cas proteins required for spacer acquisition[9-12]. In this system, the Cas1–Cas2 CRISPR genome integration complex binds to double-stranded DNA substrates (protospacers) and initiates spacer acquisition into CRISPR arrays[13]. It was thought that protospacers could only be double stranded DNA until the observation was made that some natural Cas1 proteins are directly fused to reverse transcriptase (RT) domains (RT-Cas1)[14-17], which raised the intriguing possibility of a concerted mechanism for CRISPR spacer acquisition directly from RNA.

Our solution: Transcriptional recording by CRISPR spacer acquisition from RNA

Recently my laboratory identified one RT-Cas1-containing CRISPR–Cas systems from *F. saccharivorans* capable of acquiring RNA-derived spacers heterologously in *E. coli*. Leveraging the *F. saccharivorans* CRISPR genome integration complex (*Fs*RT-Cas1–Cas2) we developed Record–seq, a

method utilizing CRISPR acquisition from RNA for transcriptional recording of molecular events within living cells (**Figure 1**) [18]. This technology allows cells to be their own biographers, constantly taking note of how their gene expression changes throughout time, which like a fossil record can be used to reconstruct an environmental history.

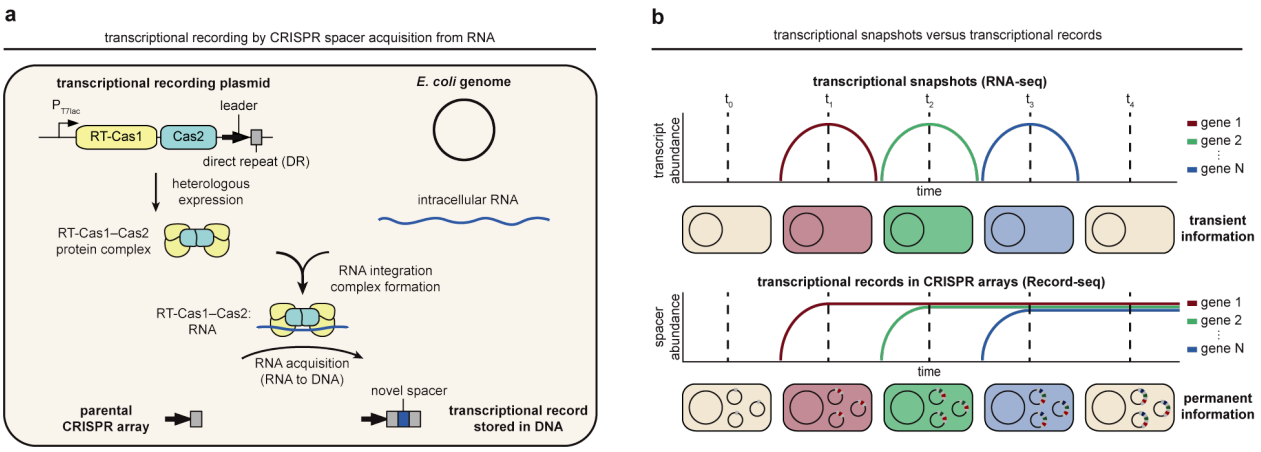

Figure 1 – Transcriptional recording by CRISPR spacer acquisition from RNA. (**a**) Expression of RT–Cas1–Cas2 leads to the acquisition of intracellular RNAs, providing a molecular memory of transcriptional events stored within DNA. **b**, Comparison of RNA-seq and Record–seq. RNA-seq captures the transcriptome of a population of cells at a single point in time, providing a transient snapshot of cellular events. By contrast, Record–seq permanently stores information about prior transcriptional events in a CRISPR array, providing a molecular record that can be used to reconstruct transcriptional events that occurred over time. [Adapted from Schmidt et al., Nature, 2018 [18]]

Over the past several years we have worked extensively to establish a suite of molecular and computational methods for efficiently retrieving and analyzing transcriptional records (**Figure 2**). The first set of challenges we had to overcome were based on CRISPR spacer acquisition being a naturally inefficient process and established sequencing-based methods were insufficient. Thus, we developed SENECA (selective amplification of expanded CRISPR arrays), a method that relies on a minimal CRISPR array consisting of a leader sequence and a single full-length DR immediately followed by a Type IIS restriction endonuclease site, which together enable Illumina adapter ligation and selective amplification of expanded, but not unexpanded (i.e., parental), CRISPR arrays. SENCA marks an improvement of several thousand-fold in detection of spacer acquisition compared to previous techniques. To analyze transcriptional records with Record–seq, we took advantage of common deep sequencing packages developed for RNA sequencing and also develop our own custom scripts and workflows, which allows us to quantify the (cumulative) gene expression of a population of cells and perform standard as well as dynamic gene expression analyses. Using Record–seq, we showed that defined stimuli, such as an invading RNA virus or the expression of inducible transgenes, as well as complex stimuli, such as oxidative or acid stress responses, result in interpretable, dose-dependent, and transcriptome-scale records of molecular events stored within the DNA of a population of cells.

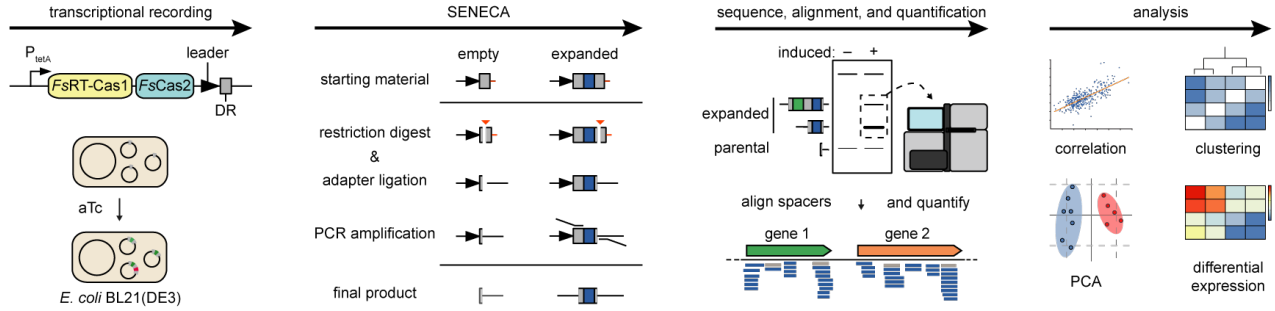

Figure 2 – Record–seq workflow. Transcriptional recording components (RT-Cas1, Cas2, and a CRISPR array composed of a leader and a direct repeat (DR)) are transformed into *E. coli* and RT-Cas1-Cas2 expression is induced to initiate transcriptional recording. Newly acquired spacers derived from RNA (i.e., transcriptional records) are retrieved using selective amplification of expanded CRISPR arrays (SENECA), purified on an agarose gel, and deep sequencing using the Illumina platform. Transcriptional records are aligned to genomic sequences using bowtie2 and quantified using featureCounts resulting in a (cumulative) gene expression matrix, which can be analyzed using standard software packages developed for RNA sequencing. [Adapted from Schmidt et al., Nature, 2018 [18]]

The future: Living diagnostics and therapeutics

All surfaces of the human body are lined with bacteria that participate in a range of normal biological as well as pathological processes[19]. To tolerate fluxes in their environment, bacteria have evolved an arsenal of natural sensors and circuits to adapt and survive. Recent efforts in synthetic biology are taking advantage of these properties and additionally engineering bacteria with gene circuits towards the goal of creating living diagnostics and therapeutics[20]. Engineered bacteria are already showing diagnostic success, including recent efforts directed at diagnosing inflammatory bowel disease, autoimmune diabetes, obesity, as well as others. However, current engineered cells rely entirely on single biosensors for the classification of pathological states, which while valuable for providing a focused assessment on a single biomarker, fail to provide a sufficiently complex picture for resolving interrelated pathological or normal biological states[20]. For example, multiple sophisticated synthetic gene circuits have been built to sense and report on gut inflammation, which is a major component of several gut and non-gut disorders, but inflammation can be masked by the consumption of cannabis and dietary supplements or mimicked by the consumption of common nonsteroidal anti-inflammatory drugs (e.g., aspirin and ibuprofen)[21, 22]. Therefore, knowledge of the inflammatory state of the gut alone is insufficient for distinguishing between even the most trivial diagnostic situations. In sum, while incredible progress in this area is being made, new tools capable of capturing more comprehensive information about cellular environments are desperately needed. In my laboratory, our goal is to overcome these challenges by developing diagnostic sentinel cells with the capacity to continuously and comprehensively monitor critical biological and pathological facets of mammalian biology via transient passage through the gastrointestinal tract.

References

1. Karczewski, K.J. and M.P. Snyder, *Integrative omics for health and disease.* Nat Rev Genet, 2018. **19**(5): p. 299-310.
2. Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics.* Nat Rev Genet, 2009. **10**(1): p. 57-63.

3. Ozsolak, F. and P.M. Milos, *RNA sequencing: advances, challenges and opportunities.* Nat Rev Genet, 2011. **12**(2): p. 87-98.
4. Schmidt, F. and R.J. Platt, *Applications of CRISPR-Cas for synthetic biology and genetic recording.* Current Opinion in Systems Biology, 2017. **5**: p. 9-15.
5. Farzadfard, F. and T.K. Lu, *Emerging applications for DNA writers and molecular recorders.* Science, 2018. **361**(6405): p. 870-875.
6. Sheth, R.U. and H.H. Wang, *DNA-based memory devices for recording cellular events.* Nat Rev Genet, 2018.
7. Hsu, P.D., E.S. Lander, and F. Zhang, *Development and applications of CRISPR-Cas9 for genome engineering.* Cell, 2014. **157**(6): p. 1262-78.
8. Doudna, J.A. and E. Charpentier, *Genome editing. The new frontier of genome engineering with CRISPR-Cas9.* Science, 2014. **346**(6213): p. 1258096.
9. Nuñez, James K., et al., *CRISPR Immunological Memory Requires a Host Factor for Specificity.* Molecular Cell, 2016. **62**(6): p. 824-833.
10. Nunez, J.K., et al., *Foreign DNA capture during CRISPR-Cas adaptive immunity.* Nature, 2015. **527**(7579): p. 535-8.
11. Nuñez, J.K., et al., *Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity.* Nat Struct Mol Biol, 2014. **21**(6): p. 528-534.
12. Yosef, I., M.G. Goren, and U. Qimron, *Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli.* Nucleic Acids Research, 2012. **40**(12): p. 5569-5576.
13. Levy, A., et al., *CRISPR adaptation biases explain preference for acquisition of foreign DNA.* Nature, 2015. **520**(7548): p. 505-510.
14. Makarova, K.S., et al., *A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action.* Biol Direct, 2006. **1**: p. 7.
15. Kojima, K.K. and M. Kanehisa, *Systematic survey for novel types of prokaryotic retroelements based on gene neighborhood and protein architecture.* Mol Biol Evol, 2008. **25**(7): p. 1395-404.
16. Simon, D.M. and S. Zimmerly, *A diversity of uncharacterized reverse transcriptases in bacteria.* Nucleic Acids Res, 2008. **36**(22): p. 7219-29.
17. Silas, S., et al., *Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein.* Science, 2016. **351**(6276): p. aad4234.
18. Schmidt, F., M.Y. Cherepkova, and R.J. Platt, *Transcriptional recording by CRISPR spacer acquisition from RNA.* Nature, 2018.
19. Schmidt, T.S.B., J. Raes, and P. Bork, *The Human Gut Microbiome: From Association to Modulation.* Cell, 2018. **172**(6): p. 1198-1215.
20. Riglar, D.T. and P.A. Silver, *Engineering bacteria for diagnostic and therapeutic applications.* Nat Rev Microbiol, 2018. **16**(4): p. 214-225.
21. Ananthakrishnan, A.N., et al., *Environmental triggers in IBD: a review of progress and evidence.* Nat Rev Gastroenterol Hepatol, 2018. **15**(1): p. 39-49.
22. Ahmed, W. and S. Katz, *Therapeutic Use of Cannabis in Inflammatory Bowel Disease.* Gastroenterol Hepatol (N Y), 2016. **12**(11): p. 668-679.