Detection and Quantitation of DNA Damage on a Genome-wide Scale Using RADAR-seq

Kelly M. Zatopek, Vladimir Potapov, Jennifer L. Ong, and Andrew F. Gardner

1New England Biolabs, Ipswich, Massachusetts, USA
2Corresponding author: kzatopek@neb.com

Published in the Molecular Biology section

The formation and persistence of DNA damage can impact biological processes such as DNA replication and transcription. To maintain genome stability and integrity, organisms rely on robust DNA damage repair pathways. Techniques to detect and locate DNA damage sites across a genome enable an understanding of the consequences of DNA damage as well as how damage is repaired, which can have key diagnostic and therapeutic implications. Importantly, advancements in technology have enabled the development of high-throughput sequencing-based DNA damage detection methods. These methods require DNA enrichment or amplification steps that limit the ability to quantitate the DNA damage sites. Further, each of these methods is typically tailored to detect only a specific type of damage. RAre DAmage and Repair (RADAR) sequencing is a DNA sequencing workflow that overcomes these limitations and enables detection and quantitation of DNA damage sites in any organism on a genome-wide scale. RADAR-seq works by replacing DNA damage sites with a patch of modified bases that can be directly detected by Pacific Biosciences Single-Molecule Real Time sequencing. Here, we present three protocols that enable detection of thymine dimers and ribonucleotides in bacterial and archaean genomes. Basic Protocol 1 enables construction of a reference genome required for RADAR-seq analyses. Basic Protocol 2 describes how to locate, quantitate, and compare thymine dimer levels in \textit{Escherichia coli} exposed to varying amounts of UV light. Basic Protocol 3 describes how to locate, quantitate, and compare ribonucleotide levels in wild-type and \textit{\Delta}RNaseH2 \textit{Thermococcus kodakarenensis}. Importantly, all three protocols provide in-depth steps for data analysis. Together they serve as proof-of-principle experiments that will allow users to adapt the protocols to locate and quantitate a wide variety of DNA damage sites in any organism. © 2022 New England Biolabs. Current Protocols published by Wiley Periodicals LLC.

**Basic Protocol 1:** Constructing a reference genome utilizing SMRT sequencing

**Basic Protocol 2:** Mapping and quantitating genomic thymine dimer formation in untreated versus UV-irradiated \textit{E. coli} using RADAR-seq

**Basic Protocol 3:** Mapping and quantitating genomic ribonucleotide incorporation in wildtype versus \textit{\Delta}RNaseH2 \textit{T. kodakarenensis} using RADAR-seq

Keywords: base excision repair • DNA damage • DNA nick • DNA repair • next-generation sequencing • SMRT sequencing
INTRODUCTION

DNA is a chemically reactive molecule and its exposure to exogenous and endogenous agents can facilitate formation of DNA damage, such as single- and double-strand breaks, alterations in nucleobases, and protein-DNA crosslinks (Chatterjee & Walker, 2017). Importantly, DNA damage can have detrimental effects on cellular fitness by leading to mutations and/or cellular apoptosis (Chatterjee & Walker, 2017). The formation and persistence of DNA damage is linked to cancer, neurological disorders, and aging (Jackson & Bartek, 2009). Therefore, it is important to understand the formation of DNA damage and how cells cope with this damage.

Across all domains of life, robust DNA repair pathways exist that are responsible for recognizing DNA damage and repairing to canonical DNA. A well-known DNA repair pathway, base excision repair (BER), consists of a cascade of enzymes that recognize and repair single-nucleobase DNA lesions and apurinic/apyrimidinic (AP) sites (Wallace, 2014). The BER cascade starts with a DNA glycosylase that recognizes specific DNA nucleobase lesions. For example, formamidopyrimidine DNA glycosylase (Fpg) recognizes 8-oxoguanine (8oxoG); uracil DNA glycosylase (UDG) recognizes deoxyuridine (dU); and T4 pyrimidine dimer glycosylase (PDG) recognizes thymine dimers. DNA glycosylases cleave the glycosidic bond between the DNA nucleobase lesion and the deoxyribose sugar, typically creating an AP site intermediate (Wallace, 2014). Following the action of a DNA glycosylase, an AP endonuclease further processes the DNA to create a 3′-OH to enable a DNA polymerase to incorporate the correct DNA nucleotide. The 5′ end is processed by either the deoxyribose phosphate lyase activity or 5′-3′ exonuclease activity of a DNA polymerase, or by the action of a flap endonuclease to create a 5′-phosphate necessary for DNA ligase to seal the nick (Wallace, 2014). There are many other DNA repair pathways, including homologous recombination, non-homologous end joining, nucleotide excision repair (NER), and mismatch repair (Huang & Zhou, 2021). Like BER, each of these pathways relies on a cascade of enzymes that recognize specific DNA damage and work together to repair the DNA.

Over the years, many assays have been developed to assess levels of DNA damage and rates of repair from a tissue or organism of interest, including the comet, TUNEL, and ELISA assays (Figueroa-Gonzalez & Perez-Plasencia, 2017). While these provide information on the specific type of DNA damage in a sample, they lack in the ability to quantitate DNA damage levels and provide site-specific location of the damage across the genome, which can be used to understand how an organism responds to different levels of a DNA damage agent or to look for DNA damage hotspots. The emergence of novel sequencing technologies has enabled the development of a variety of methods that aim to assess DNA damage levels on a genome-wide scale (Sloan, Broz, Sharbrough, & Wu, 2018).

RAre DAmage and Repair (RADAR) sequencing utilizes a reconstituted in vitro BER cascade in combination with Pacific Biosciences (PacBio) single-molecule real time (SMRT) sequencing to locate and quantitate a wide variety of DNA damage sites on a genome-wide scale (Zatopek et al., 2019). Although a wide variety of DNA damage...
Table 1  Commercially Available DNA Repair Enzymes and their Preferred Substrates

| Repair enzyme                                | Preferred substrate                  |
|----------------------------------------------|--------------------------------------|
| Uracil DNA glycosylase                        | dU                                   |
| Formamidopyrimidine DNA glycosylase           | 8oxoG, FapyA, Gh, Sp, 5-OHC           |
| T4 pyrimidine dimer glycosylase               | Thymine dimer                         |
| RNaseH2                                       | rN embedded in DNA                   |
| Alkyl adenine DNA glycosylase                 | 3mA, 7mG, eA                         |
| Endonuclease III                              | Tg, Ug, DHT, 5-OHC                   |
| Endonuclease IV                               | AP site                              |
| Endonuclease V                                | dl                                   |
| Endonuclease VIII                             | AP site, DHT, Tg, Ug                 |
| T7 endonuclease                               | Mismatches, hairpins, holiday junctions |
| Thymine DNA glycosylase                       | T:G, 5fC, 5caC                       |

sequencing methods can be utilized to map DNA damage sites genome-wide, most typically require DNA amplification or enrichment steps that prohibit quantitation of DNA damage levels (Sloan et al., 2018). Further, many methods are tailored to map only a particular DNA lesion such as ribonucleotides or thymine dimers (Ding, Taylor, Jackson, & Reijsn, 2015; Koh, Balachander, Hesselberth, & Storici, 2015; Orebaugh, Lujan, Burkholder, Clausen, & Kunkel, 2018; Hu et al., 2019). RADAR-seq utilizes PacBio SMRT sequencing, a long-read sequencing technology that sequences libraries made from unamplified genomic DNA. In addition, SMRT sequencing can readily detect N\textsuperscript{4}-methyl-cytosine (4mC) and N\textsuperscript{6}-methyl-adenine (6mA) sites and distinguish them from canonical cytosine and adenine by extending the sequencing polymerase interpulse duration time (Flusberg et al., 2010). RADAR-seq takes advantage of this by replacing DNA lesions with a patch of modified nucleotides, including 4mC and 6mA, which can then be detected by SMRT sequencing (Zatopek et al., 2019). This patch is created by an \textit{in vitro} BER cascade, which uses a DNA glycosylase or nicking enzyme to nick genomic DNA at a particular lesion site, a DNA polymerase that incorporates a patch of methylated nucleotides at the nick site, and a DNA ligase that seals the patch. Therefore, each DNA damage site is replaced with a patch of methylated nucleotides that can be detected by SMRT sequencing.

Importantly, any DNA lesion that has a corresponding DNA repair enzyme that nicks at a lesion site can be detected by RADAR-seq (Table 1). Following PacBio sequencing of RADAR-seq libraries, a downstream bioinformatics pipeline is utilized to identify patches of methylated bases present in the library. Because DNA libraries are made from unamplified genomic DNA, RADAR-seq can both locate and quantitate DNA damage levels, making it a unique DNA damage detection method. From RADAR-seq experiments, one can obtain the total number of lesions detected per million bases sequenced, as well as the genomic location and strand of each lesion.

Here, we present three protocols for locating and quantitating DNA damage in bacterial and archaeal genomes (Fig. 1). Basic Protocol 1 details the construction of a reference genome for \textit{MG1655 Escherichia coli}, which is required and utilized for RADAR-seq analysis of \textit{E. coli} in Basic Protocol 2. In this protocol, \textit{E. coli} are grown and harvested, genomic DNA is extracted and sheared into 20-kb fragments, PacBio library construction is performed on the genomic DNA, and then PacBio SMRT sequencing and microbial assembly are performed to create a polished reference genome. Importantly, the same general approach is used to create a reference genome for \textit{Thermococcus}.
Figure 1 Overview of RADAR-seq protocols. Basic Protocol 1 describes how to construct a reference genome from *E. coli* genomic DNA. Basic Protocol 2 describes how to expose *E. coli* to varying levels of UV radiation and then locate and quantitate thymine dimers in these samples. Basic Protocol 3 describes how to locate and quantitate ribonucleotides in wild-type and ΔRNaseH2 *T. kodakarensis*. All protocols describe how to perform PacBio sequencing and data analysis.

*kodakarensis*, which will be utilized in Basic Protocol 3. Basic Protocol 2 details the quantitation of thymine dimers in *E. coli* exposed to UV radiation using RADAR-seq. *E. coli* are grown and exposed to varying levels of UV radiation. The cells are harvested and genomic DNA is extracted and sheared into 2-kb fragments. These fragments are used to construct PacBio libraries and then an *in vitro* BER step is performed to convert thymine dimers to patches of methylated bases to enable detection by PacBio sequencing. The libraries are sequenced by PacBio SMRT sequencing and downstream patch detection analysis is performed to quantitate and determine the location of thymine dimers present in the *E. coli* genomic DNA. Importantly, this protocol outlines how RADAR-seq can be utilized to locate and quantitate DNA lesions (thymine dimers) in an organism (*E. coli*) exposed to a DNA-damaging agent. Basic Protocol 3 details the quantitation of ribonucleotides embedded in wild-type *T. kodakarensis* as well as *T. kodakarensis* with a deletion of a ribonucleotide excision repair enzyme (ΔRNaseH2). Wild-type and ΔRNaseH2 *T. kodakarensis* are grown and harvested. Genomic DNA is extracted, sheared into 2-kb fragments, and used for PacBio library construction. Ribonucleotides are converted into patches of methylated bases using an *in vitro* BER step, then libraries are sequenced by PacBio SMRT sequencing, and patch detection analysis is performed to quantitate and locate ribonucleotides in both strains. This protocol outlines how RADAR-seq can be utilized to locate and quantitate DNA lesions (ribonucleotides) in an organism (*T. kodakarensis*) lacking a DNA repair protein to understand the *in vivo* contribution of a DNA repair protein in genome maintenance.

Although we present the RADAR-seq workflow for detection of thymine dimers and ribonucleotides, RADAR-seq can also be utilized to map a wide variety of other DNA damage lesions (Table 1) or any process or enzyme that cleaves DNA, including CRISPR-Cas and restriction modification systems. Further, although we use bacterial and archaeal species as models for analysis, the method can be adapted for any organism of interest. RADAR-seq can be used to assess DNA damage effects of an agent, explore the persistence or rate of repair of DNA lesions, and evaluate the contributions of different repair
enzymes to *in vivo* repair of DNA lesions. In all, RADAR-seq enables a global understanding of DNA damage formation and repair and provides insight into the deleterious effects of DNA damage on cellular fitness.

**STRATEGIC PLANNING**

To perform all three protocols, genomic DNA is extracted and PacBio sequencing libraries are constructed. To create high-quality sequencing libraries, it is important to start with intact, high-molecular-weight genomic DNA. To ensure the integrity of the genomic DNA, the DNA should be analyzed using an Agilent TapeStation, an Agilent Bioanalyzer, or gel electrophoresis. Care should be taken during handling of genomic DNA, which should never be vortexed and should not be repeatedly frozen and thawed. Starting with freshly extracted genomic DNA is recommended to produce high-quality DNA libraries for high-quality sequencing output.

All three protocols utilize the PacBio Sequel instrument to produce the sequencing output. Importantly, the consumables utilized by the Sequel provide enough reagents to sequence four SMRT cells during a single run, and the reagents cannot be reused. Therefore, it is important to plan experiments to maximize utilization of consumables. Ideally, four libraries (or four pools of libraries) should be run at a time, as PacBio sequencing cells and reagent plates are set up to sequence four samples. Although we utilize the PacBio Sequel in the protocols described here, the PacBio Sequel II and Sequel IIe instruments can also be utilized to analyze RADAR-seq experiments. The Sequel II and IIe have 8-fold higher sequencing capacity compared to the Sequel and are therefore ideally suited to perform sequencing on larger genomes (>50 Mb).

Each Sequel SMRT cell can accommodate ~1,000,000 individual DNA molecules and can produce 20 Gb of sequencing data. A single SMRT cell provides enough sequencing output to create a reference genome for *E. coli* or *T. kodakarenensis*. For DNA damage quantitation experiments such as those outlined in Basic Protocol 2 and 3, up to three libraries can be pooled and sequenced in a single Sequel SMRT cell for organisms with genomes <10 Mb. If samples are pooled, unique barcodes must be ligated during library construction.

Basic Protocol 3 requires the user to grow *T. kodakarenensis* and to create a ΔRNaseH2 strain as described previously (Gehring, Sanders, & Santangelo, 2017). *T. kodakarenensis* is a hyperthermophilic, anaerobic archaeal species, and requires specialized equipment and materials for growth and strain construction. The necessary materials and equipment required for the protocol should be obtained in advance.

**CONSTRUCTING A REFERENCE GENOME UTILIZING SMRT SEQUENCING**

This protocol describes how to create a reference genome utilizing PacBio SMRT-sequencing. A reference genome is required to analyze RADAR-seq data. We demonstrate the protocol using *E. coli* to generate a reference genome for use in Basic Protocol 2. The steps include growing and harvesting *E. coli* cells, extracting and shearing *E. coli* genomic DNA, constructing a large-insert PacBio library (Fig. 2A), sequencing the library with PacBio SMRT sequencing, and utilizing PacBio software to assemble a complete genome. This protocol can be used to assemble any genome (e.g., *T. kodakarenensis* for use in Basic Protocol 3), provided the organism can be grown and genomic DNA can be extracted. Other DNA sequencing platforms such as Oxford Nanopore Technologies sequencing or Illumina sequencing (Minei, Hoshina, & Ogura, 2018) can be utilized to create a reference genome; however, we use PacBio sequencing here because RADAR-seq utilizes PacBio sequencing for DNA damage detection.
Figure 2  PacBio library construction. (A) Diagram of library construction. Extracted genomic DNA is sheared into 20-kb fragments (for genome assembly) or 2-kb fragments (for RADAR-seq). The fragments are DNA damage and end repaired, resulting in dA tails, and then PacBio SMRTbell adapters are ligated onto the fragments. (B) Agilent DNA TapeStation gel showing high-integrity extracted genomic DNA from *E. coli* (lane 2), *E. coli* genomic DNA sheared into 20-kb fragments using a Megaruptor for genome assembly to construct a reference genome (lane 3), and *E. coli* genomic DNA sheared into 2-kb fragments using Covaris sonication for RADAR-seq (lane 4). A DNA ladder is shown in lane 1.

**Materials**

- *E. coli* MG1655 (*E. coli* Genetic Stock Center, cat. no. 6300)
- LB agar plate (Sigma-Aldrich, cat. no. L5542)
- LB medium (Sigma-Aldrich, cat. no. L3522)
- Monarch Genomic DNA Extraction Kit (New England Biolabs, cat. no. T3010)
- Qubit dsDNA Broad Range (BR) and High Sensitivity (HS) Assay Kits (Thermo Fisher Scientific, cat. nos. Q32850 and Q33230)
- Genomic DNA Reagents (Agilent Technologies, cat. no. 5067-5366)
- TE light buffer (see recipe)
- AMPure PB beads (Pacific Biosciences, cat. no. 100-265-900)
- 80% (v/v) ethanol (Sigma-Aldrich, E7023) in water, freshly prepared
- SMRTbell Express Template Kit 2.0 (Pacific Biosciences, cat. no. 100-938-900)
- Sequel Binding Kit 3.0 (Pacific Biosciences, 101-626-600)
- Sequel Sequencing Kit 3.0 (Pacific Biosciences, cat. no. 101-597-900)
- Sequel SMRT Cell oil (Pacific Biosciences, cat. no. 100-621-300)
- SMRT Cells (1M v3 LR Tray, Pacific Biosciences, cat. no. 101-531-001)
- Liquid nitrogen

- 37°C incubator
- 37°C incubator shaker
- 18-ml culture tubes (VWR, cat. no. 47729-578)
- 15-ml conical tubes (VWR, cat. no. 89401-566)
Qubit Fluorometer and assay tubes (Thermo Fisher Scientific, cat. nos. Q33238 and Q32856)
Agilent 4150 or 4200 TapeStation (Agilent Technologies, cat. no. G2992AA or G2991BA)
Genomic DNA ScreenTape (Agilent Technologies, cat. no. 5067-5365)
1.5-ml DNA LoBind tubes (Eppendorf, cat. no. 022431021)
Hydro Tube (Diagenode, C30010018)
Megaruptor 2.0 (Diagenode, B06010002)
Long Hydropores (Diagenode, E07010002)
Magnetic rack (Thermo Fisher Scientific, cat. no. 12321-D)
37°C heating block (VWR, cat. no. 75838-278)
0.2-ml PCR tubes
Thermal cycler
Sample plate for PacBio sequencing (Pacific Biosciences, cat. no. 000-448-888)
Plate sealer (Thermo Scientific, cat. no. AB-1443A)
Sequel sample plate foils (Pacific Biosciences, cat. no. 100-667-700)
Pacific Biosciences Sequel Instrument (Pacific Biosciences)
Tube septa (Pacific Biosciences, cat. no. 100-667-700)
Sequel Mixing Plate (Pacific Biosciences, cat. no. 100-667-500)
Sequel Pipet Tips v2 (Pacific Biosciences, cat. no. 100-667-601)

**Grow E. coli and prepare genomic DNA**

1. Streak MG1655 *E. coli* on an LB agar plate for single colonies. Incubate overnight in a 37°C incubator.
2. Pick a single colony and inoculate 10 ml LB medium in a test tube. Incubate overnight in a 37°C incubator shaker at 250 rpm.
3. Transfer the 10-ml culture to a 15-ml conical tube and centrifuge 20 min at 2000 × g, 4°C.
4. Pour off the supernatant and extract genomic DNA using the Monarch Genomic DNA Extraction Kit according to manufacturer’s instructions.
   *Alternative genomic DNA extraction kits (e.g., Zymo Quick-DNA Microprep kit or QiaGen QIAamp DNA kit) can be utilized provided they produce high-molecular-weight genomic DNA*
5. Quantitate extracted genomic DNA using a Qubit fluorometer with Broad Range Qubit reagents.
   *The typical yield from 10 ml culture is 6-10 μg DNA.*
6. Assess the quality of the genomic DNA using an Agilent TapeStation, Genomic ScreenTape, and Genomic DNA Reagents (Fig. 2B).
   *It is important that the extracted genomic DNA be of high molecular weight. If the genomic DNA is highly fragmented (fragments < 20 kb), begin again and take care not to vortex or heat the DNA.*
7. Add 3 μg genomic DNA to 200 μl TE light buffer in a 1.5-ml DNA LoBind tube.
   *DNA LoBind tubes are important during PacBio library preparation to minimize nonspecific binding of the DNA to the tube and maximize library yield.*
8. Transfer to a Hydro Tube and then shear the genomic DNA to 20-kb fragments using a long Hydropore and Megaruptor 2.0 according to manufacturer’s instructions (Fig. 2B).
   *A Hydro Tube is required for shearing using the Megaruptor.*
9. Transfer sheared genomic DNA to a 1.5-ml DNA LoBind tube.
10. Isolate the 20-kb fragments by adding 90 μl (0.45 × volume) of AMPure PB beads and placing on the benchtop for 5 min.

11. Place the tube on a magnetic rack, allow solution to clear, and then pipette off the supernatant and save in a new DNA LoBind tube.

   *At this point, genomic DNA should remain bound to the beads, but it is beneficial to save the supernatant in case the DNA did not bind to the beads properly. To minimize sample loss, it is important to pipette the supernatant after all visible beads are stuck to the magnet.*

12. With the tube still on the magnetic rack, wash the beads twice by adding 1 ml freshly prepared 80% (v/v) ethanol and then removing it.

13. Remove tube from the magnetic rack and centrifuge for 5 s using a minicentrifuge.

14. Return tube to the magnetic rack and pipette off any residual ethanol.

15. Remove tube from the magnetic rack and place on the benchtop with the cap open for 1 min to dry off any remaining ethanol.

16. Elute genomic DNA from the beads by adding 46 μl TE light buffer and heating for 15 min in a 37°C heat block.

17. Place tube on the magnetic rack, allow solution to clear, and pipette the supernatant to a new 1.5-ml DNA LoBind tube. Discard the tube containing beads.

   *The extracted genomic DNA can be stored up to 2 years at −20°C.*

**Construct PacBio library**

The following steps require the PacBio SMRTbell Express Template Kit 2.0, which contains the necessary reagents to construct PacBio libraries, including end repair and adapter ligation reagents. This protocol is adopted from PacBio’s “Procedure and Checklist – Preparing gDNA Libraries Using the SMRTbell Express Template Preparation Kit 2.0” ([https://www.pacb.com/wp-content/uploads/Procedure-Checklist-Preparing-gDNA-Libraries-Using-the-SMRTbell-Express-Template-Preparation-Kit-2.0.pdf](https://www.pacb.com/wp-content/uploads/Procedure-Checklist-Preparing-gDNA-Libraries-Using-the-SMRTbell-Express-Template-Preparation-Kit-2.0.pdf)).

18. Make a 1:5 dilution of DNA Prep Additive by adding 1 μl to 4 μl DNA Enzyme Dilution Buffer in a 1.5-ml DNA LoBind tube.

19. Prepare a single-stranded digestion removal reaction in a 0.2-ml PCR tube as follows:

   46 μl 20-kb genomic DNA (step 17)
   7 μl DNA Prep Buffer
   1 μl NAD⁺
   1 μl diluted DNA Prep Additive (step 18)
   1 μl DNA Prep Enzyme

20. Mix gently by pipetting and place in a thermal cycler for 15 min at 37°C.

   *This step removes any large single-stranded 3’ overhangs from the 20-kb fragments.*

21. Place on ice. Add 2 μl DNA Damage Repair Mix v2, mix gently by pipetting, and place in a thermal cycler for 30 min at 37°C.

   *This step repairs DNA damage from the sheared genomic DNA, including nicks, gaps, AP sites, oxidative, UV, and deaminated bases.*

22. Place on ice. Add 3 μl End Prep Mix, mix gently by pipetting, and incubate in a thermal cycler for 10 min at 20°C, followed 30 min at 65°C.

   *This step creates DNA fragments with 5’ phosphates and 3’ -dA tails required for ligation of SMRTbell sequencing adapters the next step.*
23. Place on ice.

24. Prepare the PacBio SMRTbell adapter ligation reaction in a 0.2-ml PCR tube as follows:
   
   - 5 μl Overhang Adapter v3
   - 61 μl dA-tailed DNA fragments (step 22)
   - 1 μl Ligation Additive
   - 1 μl Ligation Enhancer
   - 30 μl Ligation Mix

   *This step ligates the SMRTbell adapters required for PacBio sequencing.*

25. Incubate in a thermal cycler for 1 hr at 20°C.

26. Isolate the DNA library as in steps 10-15, but use 44 μl AMPure beads.

27. Elute the library from the beads as in steps 16-17, but use 11 μl TE light buffer.

28. Quantitate the DNA library as in step 5.
   
   *A typical library yield for 3 μg genomic DNA input is 1000 ng (or 10 μl at 100 ng/μl).*

29. Prepare 10 μl of sample at 20 ng/μl in a new DNA LoBind tube by diluting with TE light buffer.
   
   *For example, if the library is 100 ng/μl, mix 2 μl library with 8 μl TE light to obtain 10 μl at 20 ng/μl.*

   *The completed PacBio library can be stored for ~6 months at –20°C.*

### Perform sequencing and assemble reference genome

The next steps require access to the Pacific Biosciences SMRTlink Web portal. This site is used to determine sequencing conditions, enter sequencing run parameters, initiate post-sequencing analyses (such as genome assembly), and perform data management. The steps below describe (1) “Sample Set-up”, to determine primer and sequencing polymerase binding steps, (2) “Run Design”, to enter sequencing run parameters, and (3) “SMRT Analysis”, to assemble the sequencing reads into a reference genome.

30. Go to the SMRTlink 8.0 website.
   
   *The SMRTlink address will be specific to the user’s PacBio sequencing facility.*

31. Click on “Sample Set-Up” and then on “New Calculation”, taking care to ensure that “Sequel” is selected as the desired sequencing instrument.

32. Enter the following parameters into the “New Calculation” worksheet:

   | Parameter                          | Value               |
   |-----------------------------------|---------------------|
   | Sample Comment                    | E.coli Genome Assembly (or desired sample name) |
   | Application                       | Microbial Assembly  |
   | Available Volume                  | 10                  |
   | Sample Concentration (ng/μl)      | 20                  |
   | Insert Size (bp)                  | 20000               |
   | Internal Control                  | None                |
   | Cleanup Anticipated Yield         | 50%                 |
   | Specify Concentration on Plate    | 10 pM               |
   | Cells to Bind                     | 1                   |
   | Sequencing Primer                 | Sequencing Primer v4|
   | Binding Kit                       | Sequel Binding Kit 3.0|

   *Once the parameters have been input, the software will provide directions for primer and polymerase binding.*

33. Dilute 1 μl Sequencing Primer v4 in 29 μl of 1× Elution Buffer in a 0.2-ml PCR tube. Condition the sequencing primer by heating in a thermal cycler for 2 min at 80°C, followed by a hold at 4°C.
The reagents for steps 33-46 are found in the SMRTbell Express Template Kit 2.0 and Sequel Binding Kit 3.0.

34. Bind the sequencing primer to the SMRTbell adapters by combining the following in a PCR tube and incubating for 60 min at room temperature:
   - 1.7 μl 10× Primer Buffer
   - 5.4 μl genomic DNA library (step 29)
   - 1 μl conditioned sequencing primer (step 33)

35. Dilute 1 μl Sequel Polymerase with 9 μl Sequel Binding Buffer in a PCR tube.

36. Perform a second dilution by mixing 1 μl diluted polymerase with 1 μl Sequel Binding Buffer another PCR tube.

37. In a new PCR tube, combine the following reagents and mix by pipetting:
   - 2.1 μl Sequel Binding Buffer
   - 1.7 μl ddH2O
   - 1.5 μl DTT
   - 1.5 μl Sequel dNTPs
   - 7.4 μl primer-bound DNA library (step 34)
   - 1 μl diluted Sequel polymerase (step 36)

38. Incubate in a thermal cycler for 240 min at 30°C to bind the polymerase to the primer-bound DNA libraries.

39. Add 84.8 μl Complex Dilution buffer to the tube and transfer mixture to a 1.5-ml DNA LoBind tube.

40. Add 120 μl AMPure beads and place on the benchtop for 5 min.

41. Place on the magnetic rack, allow solution to clear, and remove the supernatant.

42. Immediately add 25 μl Complex Dilution buffer and incubate on the benchtop for 15 min.

43. Place on the magnetic rack, allow to clear, and pipette the supernatant to a new DNA LoBind tube. Discard the tube containing beads.

44. Quantitate the primer/polymerase-bound library on a Qubit fluorometer using the Qubit dsDNA HS Kit.

45. Input the volume (49 μl) and the concentration (step 44) into the “New Calculations” worksheet from step 31.

   Once this is done, directions for “Final Loading Dilution” will be populated.

46. Using the directions provided in step 45, make the final loading dilution in a new 1.5-ml DNA LoBind tube.

   For example, if the volume was 49 μl and the concentration was 0.5 ng/μl, you would be instructed to combine 53.4 μl Sequel Complex Dilution Buffer, 22.1 μl prepared sample (step 43), 8.5 μl DTT, and 1.0 μl Sequel Additive.

47. Place a Sequel sample plate on ice and pipette the sample (step 46) to well A1 of the sample plate.

48. Remove the Sequel sequencing kit from the freezer and place on the benchtop to thaw (∼1 hr).

49. Turn on the plate sealer and allow to come to temperature (∼15 min).

50. Seal the plate (step 47) using the plate sealer and Sequel sample plate foil.
51. On the SMRTlink Web portal, click on “Run Design” then “Create New Design”, ensuring that the Sequel instrument is selected.

52. On the left-hand side, input the Run Name (e.g., “Ecoli Genome Assembly-DDMMYYYY”).

53. Under sample information, input the following parameters, then hit “Save”:

| Parameter                     | Value                      |
|-------------------------------|----------------------------|
| Application                   | Microbial Assembly         |
| Well Sample Name              | Ecoli Genome Assembly (or desired ID) |
| Sample Well                   | A01                        |
| Template Prep Kit             | SMRTbell Express Template Prep Kit 2.0 |
| Binding Kit                   | Sequel Binding Kit 3.0     |
| Sequencing Kit                | Sequel Sequencing Plate 3.0 (4rxn) |
| DNA Control Complex:          | Leave blank                |
| Insert Size (bp)              | 20000                      |
| On-plate Loading Concentration| 10 pM                      |
| Movie Time per SMRT-cell      | 20 hr                      |
| Use Pre-extension             | Yes                        |
| Pre-Extension Time            | 2 hr                       |

54. At the PacBio Sequel instrument, click “Select Existing Run”, then select the run name designated above (“Ecoli Genome Assembly-DDMMYYYY”) and click “Continue”.

55. Remove cap from the Sequel sequencing oil tube and place a tube septum onto the tube.

56. Unlock the instrument door by selecting the “Locked” button in the bottom right-hand corner of the Sequel screen, then pull open the instrument door.

57. Load the following into the Sequel Instrument:
   - sample plate (step 50)
   - mixing plate
   - sequencing oil (step 55)
   - sequencing reagent plate (step 48)
   - Sequel SMRT Cells
   - three boxes of sequencing pipettes

58. Click “Load” on the Sequel screen, followed by “Scan”, and “Empty”. Ensure the Sequel trashcan (in the black bucket in the instrument) is empty. If it is not, empty contents into a sharps bucket and place back in the instrument.

59. Ensure that there is an appropriate amount of liquid nitrogen to run the instrument. Once confirmed, click the “Check Nitrogen” button.

   *The Sequel instrument requires liquid nitrogen to prevent oxygen-induced DNA damage. For a large liquid nitrogen dewar, >30% liquid nitrogen is sufficient.*

60. Click “Start” at the top right portion of the screen to initiate the sequencing run.

   *A 20-hr sequencing run takes ~24.5 hr from start to finish due to instrument warm-up and downtime.*

61. Once the run is done, go to the SMRTlink Web portal and click “SMRT Analysis” and “Create New Analysis”.

62. Input the following:

| Parameter                     | Value                      |
|-------------------------------|----------------------------|
| Analysis Name                 | “Ecoli Genome Assembly”    |
| Data Type                     | “Continuous Long Reads”    |
| Sequencing Data Set           | “Ecoli Genome Assembly-DDMMYYYY” |
63. Click “Next” and then set the following:

- **Analysis Application** “Microbial Assembly”
- **Genome Length** 4700000

64. Click “Start” to initiate assembly of the *E. coli* genome.

65. Once the assembly is finished, download the assembled genome by going to “SMRT Analysis” in the SMRTlink Web portal, clicking on “Ecoli Genome Assembly”, “Data”, and “Final Polished Assembly”.

   *A single polished contig will be obtained in FASTA format.*

### MAPPING AND QUANTITATING GENOMIC THYMINE DIMER FORMATION IN UNTREATED VERSUS UV-IRRADIATED *E. coli* USING RADAR-seq

Here, the user will use RADAR-seq to locate and quantitate thymine dimers in bacteria (Fig. 3A). To exemplify the approach, MG1655 *E. coli* cells are exposed to varying levels of UV radiation and then RADAR-seq DNA libraries are constructed to allow detection and quantitation of thymine dimer formation catalyzed by UV radiation on a genome-wide scale. First, *E. coli* cells are grown and then exposed to varying levels of UV radiation. Second, genomic DNA is extracted, sheared, and used to construct PacBio libraries. Third, *in vitro* BER steps are performed to convert thymine dimers into patches of modified nucleotides. Finally, libraries are sequenced by PacBio SMRT sequencing and downstream patch detection analysis is carried out to quantitate and map modified patch sites, where each patch corresponds to a single thymine dimer lesion. Once completed, the user will have quantitated the number of thymine dimers detected per million bases sequenced and know the genomic location of each thymine dimer. Here, we utilize RADAR-seq to quantitate thymine dimers in *E. coli* genomic DNA, but the same protocol can be followed for any tissue or organism that can be exposed to UV radiation in a controlled environment and from which high-quality genomic DNA can be extracted. The steps outlined are for one experimental replicate, but a total of three experimental replicates should be performed to measure variability.

**Additional Materials** *(also see Basic Protocol 1)*

- 5 × M9 minimal-salts bacterial growth medium (Sigma-Aldrich, cat. no. M6030)
- PreCR Enzyme Mix lacking T4 PDG (available upon request from New England Biolabs)
- PacBio Barcoded Adapters (Pacific Biosciences, cat. no. 101-628-400, 101-628-500, or 102-009-200)
- 10× rCutSmart Buffer (New England Biolabs, cat. no. B6004)
- Exonucleases III and VII (New England Biolabs, cat. nos. M0206 and M0379)
- 10× ThermoPol Buffer (New England Biolabs, cat. no. B9004)
- 1 mM methylated dNTP mix (see recipe)
- 50 mM NAD⁺ (New England Biolabs, cat. no. B9007)
- T4 pyrimidine dimer glycosylase (PDG, New England Biolabs, cat. no. M0308)
- *Tth* Endonuclease IV (New England Biolabs, cat. no. M0294)
- *Bst* DNA polymerase, full length (New England Biolabs, cat. no. M0328)
- *Taq* DNA ligase (New England Biolabs, cat. no. M0208)
- *E. coli* reference genome (see Basic Protocol 1)
- 500-ml Erlenmeyer flask (VWR, cat. no. 10536-926)
- 250-ml centrifuge bottle (VWR, cat. no. 47735-688)
- 150 × 15-mm Petri dishes (Fisher Scientific, cat. no. FB0875714)
- Spectrolinker X-1000 (Spectro-UV)
Figure 3  Genome-wide thymine dimer detection using RADAR-seq. (A) Thymine dimers in PacBio libraries (created from genomic DNA purified from *E. coli* exposed to UV radiation) were replaced with a patch of methylated nucleotides using (1) T4 pyrimidine dimer glycosylase (PDG) to remove the thymine dimer lesion, (2) *Tth* endonuclease IV to nick the DNA backbone, (3) gap repair using full-length *Bst* DNA polymerase together with methylated dNTP (*mdNTP*), and (4) gap closure by ligation using *Taq* DNA ligase and NAD⁺, as described in Basic Protocol 2. (B) Bar graph displaying the average number of thymine dimers detected (per million bases sequenced) by RADAR-seq in *E. coli* cells exposed to UV light (254 nm). Error bars indicate standard deviation for at least three biological replicates. (C, D) Location of thymine dimers on a genome-wide scale for no UV exposure or UV exposure at 200 J/m². Circular plots were generated using Circos software (Krzywinski et al., 2009). Red and blue line plots correspond to frequency of observed patches (per million bases sequenced) on the top and bottom strand, respectively. Frequency of patches was computed using a rolling average with a 10-kb window size. Figure adapted from Zatopek et al. (2019).

50-ml conical tubes (VWR, cat. no. 89401-562)
Covaris miniTUBEs, clear (Covaris, cat. no. 520064)
Covaris ME220 Focused-ultrasonicator (Covaris, cat. no. 500506)
Tools and scripts for data analysis (available at [https://github.com/potapovneb/CP-RADAR-seq;10.5281/zenodo.6914938](https://github.com/potapovneb/CP-RADAR-seq;10.5281/zenodo.6914938))

Grow *E. coli* and expose to UV
1. Streak MG1655 *E. coli* on an LB agar plate for single colonies. Incubate overnight in a 37°C incubator.
2. Pick a single colony and inoculate 250 ml LB medium in a 500-ml Erlenmeyer flask. Incubate overnight in a 37°C incubator shaker at 250 rpm until an OD$_{600}$ of 0.8 is reached.

3. Transfer culture to a 250-ml centrifuge bottle and pellet cells by centrifuging 20 min at 10,000 $\times$ g, 4°C.

4. Pour off the supernatant and resuspend cell pellet in 250 ml of 1× M9 minimal medium.

*M9 minimal medium is sold as a 5× solution and should be diluted to 1× using ddH$_2$O in a sterile bottle.*

5. Pipette 15 ml resuspended *E. coli* onto an empty 150 × 15-mm Petri dish.

6. Place the dish in a Spectrolinker XL-1000 and expose cells to 10 J/m$^2$ UV radiation at 254 nm.

*Choose the “Energy” option on the Spectrolinker XL-1000 keypad. The units on the instrument are in $\times 100$ μJ/cm$^2$. Therefore, to deliver 10 J/m$^2$ UV radiation, type “10” on the keypad followed by “Enter”.*

7. Pipette the 15-ml cell suspension into a 50-ml conical tube and place on ice.

8. Repeat steps 5-7 on four additional 15-ml aliquots of culture, exposing them to increasing levels of radiation at 25, 50, 100, and 200 J/m$^2$.

*Use a new Petri dish and serological pipette for each sample.*

9. Remove another 15-ml aliquot of culture and place in a 50-ml conical tube on ice to serve as the no-exposure control.

*There are now six 15-ml aliquots of cells on ice ranging from 0 to 200 J/m$^2$ exposure.*

10. Pellet all samples by centrifuging 20 min at 2000 $\times$ g, 4°C.

11. Remove M9 medium supernatant from each tube and discard.

12. Proceed immediately to genomic DNA extraction or store samples in a −20°C freezer until use.

*Cells can be stored for up to 6 months at −20°C*

**Extract and shear genomic DNA**

13. Extract, quantitate, and assess the integrity of genomic DNA from UV-irradiated cells as described (see Basic Protocol 4-6).

*The typical yield from each 15-ml aliquot of cells is 6-8 μg genomic DNA.*

14. Add 2 μg genomic DNA to 200 μl TE light buffer and pipette into clear Covaris miniTUBEs.

*Take care to avoid air bubbles at the bottom of the Covaris tube, as the presence of air bubbles can negatively affect DNA shearing.*

15. Shear genomic DNA into 2-kb fragments using the Covaris ME220 Focused-ultrasonicator with the following settings:

| Treatment | 900.0 s |
| Peak Power | 10.0 |
| Duty Factor | 20.0% |
| Cycles Burst | 1000 |
| Average Power | 2.0 |
Do not shear genomic DNA in water, as this facilitates oxidative DNA damage (Chen, Liu, Evans, & Ettwiller, 2017).

16. Pipette sheared genomic DNA into a 1.5-ml DNA LoBind tube. Then, centrifuge the miniTUBE for 20 s in a minicentrifuge to remove additional sample.

   *It is rare to recover the full 200 μl from the Covaris miniTUBE.*

17. Isolate the 2-kb DNA fragments as described (see Basic Protocol 1, steps 10-17), but use 0.6 × volume of AMPure beads in step 10, and use 47 μl TE light buffer to elute the DNA in step 16.

   *For example, 200 μl sample × 0.6 = 120 μl AMPure beads.*

18. Quantitate the genomic DNA and confirm that the DNA is sheared into 2-kb fragments (see Basic Protocol 1, steps 5-6; Fig. 2).

**Construct PacBio libraries**

19. Construct PacBio libraries from all DNA samples as described (see Basic Protocol 1, steps 18-25) with the following modifications:
   a. During DNA repair (step 21), use 2 μl PreCR lacking T4 PDG in place of DNA Damage Repair Mix v2.
   b. During DNA ligation (step 24), use 5 μl PacBio Barcoded Adapter instead of Overhang Adapter v3.0.

   *In this way, each sample (i.e., DNA from each UV exposure level) gets a unique SMRTbell barcoded adapter that allows for sample pooling during PacBio sequencing.*

20. Isolate libraries as described (see Basic Protocol 1, steps 10-17), but use 59 μl AMPure beads in each tube, and use 44 μl TE light buffer to elute the DNA.

21. Add 5 μl of 10× rCutSmart Buffer to each tube, followed by 0.5 μl Exonuclease III and 0.5 μl Exonuclease VII. Mix by pipetting and incubate 30 min at 37°C.

22. Place on ice and remove the exonucleases by repeating the AMPure cleanup using 30 μl beads per tube and 21 μl TE light buffer for elution.

23. Quantitate each library as described (see Basic Protocol 1, step 5).

   *Starting with 2 μg genomic DNA, a typical library yield is ~500 ng DNA.*

24. Proceed immediately to the next step to avoid introduction of nicks into the genomic DNA.

   *These nicks can arise spontaneously or as a result of freeze-thaw cycles.*

**Perform RADAR-seq BER for thymine dimer replacement**

25. Bring 500 ng of each library to a final volume of 20 μl using dH2O.

26. To nick the DNA at thymine dimer sites, combine the following in a PCR tube for each library:

   - 20 μl DNA library
   - 5 μl 10× ThermoPol Buffer
   - 5 μl methylated dNTP mix
   - 1 μl 50 mM NAD⁺
   - 16.5 μl ddH₂O
   - 1 μl T4 PDG

27. Mix well by pipetting and incubate 1 hr at 37°C.

28. While the nicking reactions are incubating, prepare the following master enzyme mix and mix well by pipetting:
3 μl *Tth* Endonuclease IV
3 μl full-length *Bst* DNA polymerase
6 μl *Taq* DNA ligase

29. When the incubation is complete, add 1.5 μl enzyme mix (step 28) to each nicked sample, mix well by pipetting, and incubate 25 min at 55°C.

   *In this step, a patch of methylated bases is incorporated and ligated into each DNA library, replacing the excised thymine dimers.*

30. Place reactions on ice and isolate libraries as described, using 30 μl AMPure beads and 44 μl TE light buffer to elute the DNA.

31. Treat with exonucleases as described in steps 21-23, but use 11 μl TE light buffer to elute the library from the beads.

   *Samples are now ready for PacBio SMRT sequencing. Ideally, samples will be >10 ng/μl, but samples as low as 3 ng/μl can be sequenced. If a large sample loss occurs during RADAR-seq BER and DNA is not recovered, see Troubleshooting.*

   *RADAR-seq libraries can be stored for up to 6 months at −20°C.*

**Pool samples and perform setup and sequencing**

The next steps describe how to pool samples for PacBio sequencing and then modify the sequencing protocol from Basic Protocol 1 (steps 30-60) for this protocol. The six libraries are pooled into two samples to limit the sequencing cost. For more information on pooling of samples, see Strategic Planning.

32. Pool the libraries into two samples, each containing three libraries, with a final volume of 1-15 μl and a pool concentration of 5-20 ng/μl.

   *For example, if the libraries are each at 10 ng/μl, combine 4 μl each from the 0, 10, and 25 J/m² samples to make pool 1, and combine 4 μl each from the 50, 100, 200 J/m² samples to make pool 2. This will give two 12-μl pools at 10 ng/μl. The actual pooling will vary depending on the library concentrations determined in step 31.*

33. Perform “Sample Set-Up” calculation for both library pools as described (see Basic Protocol 1, steps 30-32), but enter the following parameters into the “New Calculation” worksheet:

   | Sample Comment       | UV DNA Pool 1          |
   |----------------------|------------------------|
   | Application          | Continuous Long Reads  |
   | Available Volume (μl)| 12 (may differ based on pool created) |
   | Sample Concentration (ng/μl) | 10 (may differ based on pool created) |
   | Insert Size (bp)     | 2000                   |
   | Internal Control     | None                   |
   | Cleanup Anticipated Yield | 50%                    |
   | Specify Concentration on Plate | 10 pM                |
   | Cells to Bind        | 1                      |
   | Sequencing Primer    | Sequencing Primer v4   |
   | Binding Kit          | Sequel Binding Kit 3.0 |

   *It is critical to use these parameters and not those in Basic Protocol 1, as the directions provided from this calculation will differ from the previous calculation.*

34. Proceed with sample preparation for both pools (see Basic Protocol 1, steps 33-50) using the directions provided by the Sample Set-Up calculator (step 33 above) to bind the PacBio sequencing primer and polymerase.

   *The first UV DNA pool will go into well A1 of the sample plate and the second pool will go into well B1.*
35. Perform “Run Design” and input sample information for the first pool as described (see Basic Protocol 1, steps 51-53) with the following information:

a. Use the Run Name “Ecoli UV RADAR -DDMMYYY” in step 52.
b. Use the following parameters in step 53:

| Parameter                      | Value                                      |
|--------------------------------|--------------------------------------------|
| Application                    | Continuous Long Reads                      |
| Well Sample Name               | UV DNA Pool 1                              |
| Sample Well                    | A01                                        |
| Template Prep Kit              | SMRTbell Express Template Prep Kit 2.0     |
| Binding Kit                    | Sequel Binding Kit 3.0                     |
| Sequencing Kit                 | Sequel Sequencing Plate 3.0 (4rxn)         |
| DNA Control Complex            | Leave blank                                |
| Insert Size (bp)               | 2000                                       |
| On-plate Loading Concentration | 10 pM                                      |
| Movie Time per SMRT-cell       | 20 hr                                      |
| Use Pre-extension              | Yes                                        |
| Pre-Extension Time             | 1 hr                                       |
| Barcoded Sample Options        | Yes (from pop-up input barcodes)           |

36. Click “Add Sample” and then input information for UV DNA Pool 2. Save the run information.

37. Sequence the libraries (see Basic Protocol 1, steps 54-60), making sure to choose “Ecoli UV RADAR -DDMMYYYY” as the run name in step 54.

**Perform patch detection analysis**

Following sequencing, the sequencing data and the following set of scripts are used to extract "patches" of modified bases from each DNA library. Details regarding the initial methodology of patch detection can be found in Zatopek et al. (2019). A detailed description of the RADAR-seq computational workflow, with all necessary custom processing scripts, is available on GitHub at https://github.com/potapovneb/CP-RADAR-seq;10.5281/zenodo.6914938. Below, we provide a summary of the steps involved and advise users to follow the workflow on GitHub.

38. Demultiplex the DNA pools to create individual BAM subreads for each DNA library.

   *This produces a PacBio BAM subreads file containing sequencing reads for a given barcode.*

39. Align sequencing reads to the *E. coli* reference genome using the PacBio pbmm2 tool.

   *This produces a BAM file with aligned sequencing reads.*

40. Perform single modification detection analysis on the aligned reads using the PacBio ipdSummary tool and the custom scripts.

   *This produces a set of single-molecule PacBio CSV modification detection files organized into a set of directories. To speed up computations, sets of non-overlapping single molecules are processed together.*

41. Process the modification detection files using custom scripts to identify stretches of modified bases (patches) in the sequencing data.

   *This produces a table providing the genomic strand, the start and stop locations of the detected patch, and the DNA sequence of the detected patch. An example RADAR-seq output table ‘radarseq_patches.csv’ is available on Zenodo in the ‘example/’ folder.*
Here, wildtype and a ribonucleotide excision repair (ΔRNaseH2) mutant of the archaeon *T. kodakarensis* are grown and then RADAR-seq DNA libraries are prepared to detect and quantitate ribonucleotides embedded in their genomic DNA on a genome-wide scale (Fig. 4A). During DNA replication, a DNA polymerase must choose the correct dNTP over the corresponding rNTP. Because the intracellular concentration of rNTPs is typically orders of magnitude higher than dNTPs (McElhinny et al., 2010), DNA polymerases often spuriously incorporate rNTPs, and their persistence in the genome leads to genome instability (Joyce, 1997). Importantly, enzymes such as RNaseH1 and RNaseH2 initiate the ribonucleotide excision repair (RER) pathway to remove rNTPs from the genome (Kochiwa, Tomita, & Kanai, 2007; Kellner & Luke, 2020). To understand the...
contribution of RNaseH2 to ribonucleotide repair in *T. kodakarensis*, RADAR-seq is used here to locate and quantitate rNTPs in wildtype and ΔRNaseH2 strains. First, the ΔRNaseH2 strain must be created by the user. This is detailed in Gehring et al. (2017) and is not included in the steps below. Next, wildtype and ΔRNaseH2 *T. kodakarensis* are grown and genomic DNA is extracted from both strains. PacBio libraries are then constructed and *in vitro* BER steps are performed to convert rNTPs into patches of modified nucleotides. Finally, the libraries are then sequenced by PacBio SMRT sequencing, and downstream analysis is performed on the sequencing output to detect patches of modified bases. Once completed, the user will have quantitated the number of rNTPs detected per million bases sequenced and know the genomic location of each rNTP. Here, we utilize RADAR-seq to quantitate rNTPs in *T. kodakarensis* genomic DNA, but the same protocol can be followed for any organism in which genetics are available for strain construction and high-quality genomic DNA can be extracted. The steps outlined are for one experimental replicate, but a total of three experimental replicates should be performed to measure variability.

**Additional Materials (also see Basic Protocols 1 and 2)**

Wildtype *T. kodakarensis* (ATCC, cat. no. BAA-918)
ΔRNaseH2 *T. kodakarensis* (Gehring et al., 2017)
Wildtype and ΔRNaseH2 *T. kodakarensis* reference genomes (see Basic Protocol 1)
RNaseH2 (New England Biolabs, cat. no. M0288)

**Grow cells, extract genomic DNA, and construct libraries**

1. Grow the wildtype and ΔRNaseH2 *T. kodakarensis* strains as described in Gehring et al. (2017).
2. Extract genomic DNA from both cultures using the NEB Monarch Genomic DNA extraction kit according to the manufacturer’s instructions for Gram-negative bacteria.
3. Shear genomic DNA into 2-kb fragments as described (see Basic Protocol 2, steps 14-18).
4. Construct PacBio libraries as described (see Basic Protocol 1, steps 18-25), but use 5 μl PacBio Barcoded Adapter instead of Overhang Adapter v3.0 in step 24.
   
   In this way, the two samples will have unique SMRTbell barcoded adapters to allow for sample pooling during sequencing.
5. Remove any unligated/nick-containing PacBio fragments from the libraries as described (see Basic Protocol 2, steps 20-24).

**Perform RADAR-seq BER for ribonucleotide replacement**

6. Bring 500 ng of each library to a volume of 20 μl using ddH₂O.
7. To replace ribonucleotides with a patch of modified nucleotides, combine the following in a PCR tube:
   
   20 μl DNA library
   5 μl 10× ThermoPol Buffer
   5 μl methylated dNTP mix
   1 μl 50 mM NAD⁺
   16.5 μl ddH₂O

8. Prepare the following enzyme master mix and mix well by pipetting:
   
   2 μl RNaseH2
   1 μl full-length Bst DNA polymerase
   2 μl Taq DNA ligase
9. Add 2.5 μl enzyme mix to each reaction in step 7, mix well by pipetting, and incubate 5 min at 37°C, followed by 25 min at 55°C.

   *In this step, ribonucleotides are replaced with a patch of methylated bases that are then ligated into the DNA fragments.*

10. Clean up libraries as described (see Basic Protocol 1, steps 10-17), but use 30 μl AMPure beads in step 10 and use 44 μl TE light buffer to elute the libraries in step 16.

11. Treat with exonucleases as described (see Basic Protocol 2, steps 21-22), but use 11 μl TE light buffer to elute the libraries from the beads.

12. Quantitate each library as described (see Basic Protocol 1, step 5).

**Pool samples and perform setup and sequencing**

The next steps describe how to pool the two libraries into one sample for PacBio sequencing and then modify the sequencing protocol from Basic Protocol 1 (steps 30-60).

13. Pool the libraries into a single sample with a final volume of 5-15 μl and pool concentration of 5-20 ng/μl.

   *For example, for WT and ΔRNaseH2 libraries at 10 ng/μl, pool 4 μl of each library to give 8 μl at a final pool concentration of 10 ng/μl. The actual pooling will vary depending on the library concentrations determined in step 12.*

14. Perform “Sample Set-Up” calculation for the pooled library as described (see Basic Protocol 1, steps 30-32), but enter the following parameters into the “New Calculation” worksheet:

   | Sample Comment       | T. kodakarensis RNaseH2 Pool |
   |----------------------|------------------------------|
   | Application          | Continuous Long Reads       |
   | Available Volume (μl)| 8 (may differ based on pool created) |
   | Sample Concentration (ng/μl) | 10 (may differ based on pool created) |
   | Insert Size (bp)     | 2000                        |
   | Internal Control     | None                        |
   | Cleanup Anticipated Yield | 50%                         |
   | Specify Concentration on Plate | 10 pM                      |
   | Cells to Bind        | 1                           |
   | Sequencing Primer    | Sequencing Primer v4        |
   | Binding Kit          | Sequel Binding Kit 3.0      |

   *It is critical to use these parameters and not those in Basic Protocol 1, as the directions provided from this calculation will differ from the previous calculation.*

15. Proceed with sample preparation (see Basic Protocol 1, steps 33-50) using the directions provided by the Sample Set-Up calculator (step 14 above) to bind the PacBio sequencing primer and polymerase.

16. Set up the sequencing run as described (see Basic Protocol 1, steps 51-60) with the following modifications:

   a. Use the Run Name “T. kodakarensis RNaseH2 RADAR DDMMYYYYY” in steps 52-54.

   b. Use the following parameters in step 53:

   | Application          | Continuous Long Reads       |
   | Well Sample Name     | T. kodakarensis RNaseH2     |
   | Sample Well          | A01                         |
   | Template Prep Kit    | SMRTbell Express Template Prep Kit 2.0 |
   | Binding Kit          | Sequel Binding Kit 3.0      |
   | Sequencing Kit       | Sequel Sequencing Plate 3.0 (4rxn) |
DNA Control Complex  Leave blank
Insert Size (bp)  2000
On-plate Loading Concentration  10 pM
Movie Time per SMRT-cell  20 hr
Use Pre-extension  Yes
Pre-Extension Time  1 hr
Barcoded Sample Options  Yes (from pop-up input barcodes)

17. Save the run. Sequence the libraries (see Basic Protocol 2, steps 54-60), being sure to choose “T. kodakarensis RNaseH2 RADAR DDMMYYYY” as the run name in step 54.

18. Perform patch detection analysis as described (see Basic Protocol 2, steps 37-40) to obtain a table providing the genomic strand, the start and stop locations of the detected patch, and the DNA sequence of the detected patch.

REAGENTS AND SOLUTIONS

**Methylated dNTP mix, 1 mM**

- 1 μl 100 mM dTTP (New England Biolabs, cat. no. N0446)
- 1 μl 100 mM dGTP (New England Biolabs, cat. no. N0446)
- 1 μl 100 mM N⁴-methyl-dCTP (Trilink, cat. no. N2057-1)
- 1 μl 100 mM N⁶-methyl-dATP (Trilink, cat. no. N2025-1)
- 96 μl PCR-grade water

Combine in a DNA LoBind Tube (Eppendorf, cat. no. 022431021)
Store up to 6 months at –20°C

**TE light buffer**

- 400 μl 1 M Tris-HCl, pH 7.5 (final 10 mM)
- 8 μl 0.5 M EDTA (final 0.1 mM)

Bring to 40 ml with PCR-quality water
Store up to 6 months at room temperature

COMMENTARY

**Background Information**

The formation and persistence of DNA damage can have detrimental effects on cellular fitness (Chatterjee & Walker, 2017). Understanding how DNA damage forms, how it is repaired, and how it can alter biological processes has been an active area of research since DNA damage and repair mechanisms were first discovered over 90 years ago (reviewed in Friedberg, 2008). Genetic and biochemical approaches have revealed the presence of many complex enzymatic pathways that recognize DNA damage and repair it back to canonical DNA. Multiple methods have been developed to detect the presence of DNA damage from an organism of interest, including the well-established comet assay (Ostling & Johanson, 1984). While this method can detect the presence of DNA damage in a pooled sample, it lacks the ability to site-specifically locate the genomic position of DNA damage and quantify levels of damage at a single-molecule level. More recently, high-throughput (HT) sequencing-based methods such as Ribose-seq, Excision-seq, and CPD-seq have been used to locate DNA damage on a genome-wide scale; consequently, RADAR-seq was developed to locate and quantitate a wide variety of DNA damage lesions on a genome-wide scale (Bryan, Ransom, Adane, York, & Hesselberth, 2014; Koh et al., 2015; Mao, Smerdon, Roberts, & Wyrick, 2016; reviewed in Sloan et al., 2018; Mingard, Wu, Mckeague, & Sturla, 2020). RADAR-seq has several advantages compared to other HT sequencing-based DNA damage detection methods. Whereas most methods are tailored to locate a specific DNA lesion (e.g., ribonucleotides or deoxyuridines), RADAR-seq can detect a wide variety of DNA lesions (Table 1) by simply pairing a repair enzyme with the desired lesion during the in vitro BER step, as shown here for thymine dimers and ribonucleotides. This enables the user to directly compare levels of different DNA lesions in a sample. Importantly, any nick present within the
extracted DNA (whether produced enzymatically or non-enzymatically) can be detected by RADAR-seq. For example, RADAR-seq can be tailored to look at any enzyme or process that cleaves the phosphodiester backbone, such as Cas nucleases or restriction modification enzymes, making RADAR-seq one of the most widely applicable HT sequencing-based DNA damage detection methods. Further, most HT sequencing-based DNA damage detection methods require an enrichment and/or amplification step, which can lead to sample loss or limit the ability to quantitate the abundance of a particular DNA lesion (Bryan et al., 2014; Clausen et al., 2015; Ding et al., 2015; Hu et al., 2019; Keszthelyi, Daigaku, Ptasinska, Miyabe, & Carr, 2015; Koh et al., 2015). Because RADAR-seq libraries do not require enrichment or DNA amplification, and both damaged and undamaged DNAs are sequenced, production of DNA lesions can be quantitated by RADAR-seq analysis. Moreover, many DNA damage detection methods require the presence of abundant DNA lesions in the genomic sample to create DNA sequencing libraries, as cleaving at the DNA lesion site creates the fragments necessary for short-read HT-sequencing (Bryan et al., 2014; Hu et al., 2019). Because RADAR-seq does not rely on the presence of a DNA lesion to create fragments for sequencing, but rather uses Covaris shearing to fragment the DNA, rare DNA damage events (as few as 1 per million bases sequenced) can be detected by RADAR-seq (Zatopek et al., 2019). RADAR-seq library construction is straightforward and combines the traditional PacBio library construction workflow with a BER step. Finally, data analysis utilizes readily available scripts that provide a simplified format for data analysis.

Despite its many advantages, there are several limitations of RADAR-seq that should be mentioned. First, RADAR-seq utilizes PacBio sequencing, which is typically more costly than alternative sequencing methods like Illumina or ONT, and may not be available to all users, although several sequencing cores currently offer PacBio sequencing services. Further, because PacBio libraries are constructed from unamplified genomic DNA, obtaining high genome coverage can be difficult, especially for larger genomes. RADAR-seq requires at least three-fold coverage of each fragment to be analyzed. While this seems like a low threshold, it can be difficult to accomplish with very large genomes unless multiple SMRT Cells are utilized. Second, the ability for RADAR-seq to obtain single-nucleotide resolution of DNA damage/nick sites is dependent on the sequence context in which the damage/nick occurs. Because the RADAR-seq BER step utilizes a dNTP pool containing methylated dA and dC and canonical dT and dG, the accuracy of DNA nick/lesion detection is lower if the lesion is present in a GT-rich region. Finally, as with all HT sequencing-based methods, the accuracy of the output depends highly on how well the library construction process goes. There are many enzymatic steps that must be accomplished for successful library construction and RADAR-seq, and loss of enzyme activity in the workflow can affect sequencing output results. For example, during the BER step, it is critical to have enough Taq ligase that all patch sites are ligated. A loss of Taq ligase activity will lead to loss of fragments containing DNA damage sites and inaccurate results.

Basic Protocol 1 describes how to construct a bacterial reference genome using PacBio SMRT sequencing. The creation of a reference genome is essential for mapping and analysis of RADAR-seq data. In this protocol, PacBio Sequel SMRT sequencing was used to sequence the genomic DNA, and the tools provided in the SMRTlink Web portal were utilized to assemble the reference genome. While other sequencing platforms such as Illumina or ONT can be utilized for genome assembly (Minei et al., 2018), they require downstream user-driven assembly using a variety of bioinformatics tools. PacBio provides an easy-to-navigate user interface, the SMRTlink Web portal, which enables a quick and simple method of genome assembly.

Basic Protocol 2 describes how to utilize RADAR-seq to locate and quantitate thymine dimers by creating a patch of modified bases at each thymine dimer site using a cascade of BER enzymes, followed by PacBio sequencing to generate sequencing output. An alternative method to thymine dimer detection is XR-seq (Hu, Adar, Selby, Lieb, & Sancar, 2015), which captures small DNA fragments produced by the organism’s NER repair pathway and an anti–cyclobutane pyrimidine dimer (CPD) antibody. The captured fragments are amplified and sequenced to determine the location of the thymine dimer. Importantly, XR-seq relies on the presence and efficiency of the in vivo NER cascade and the anti-CPD antibody for DNA capture. Because XR-seq enriches sites that contain CPDs, it cannot be used to quantitate levels of thymine dimers.
Basic Protocol 3 describes how to utilize RADAR-seq to locate and quantitate ribonucleotides. Ribonucleotides are replaced with a patch of modified bases using a BER cascade and PacBio sequencing to generate output. There are other HT sequencing-based ribonucleotide detection methods, including HydEn-seq and Ribose-seq (Clausen et al., 2015; Koh et al., 2015). In HydEn-seq, genomic DNA containing ribonucleotides is fragmented with KOH, followed by a series of enzymatic steps including PCR amplification to create libraries. In Ribose-seq, genomic DNA containing ribonucleotides is cleaved by RNaseH2, followed by a series of enzymatic steps including DNA circularization followed by PCR amplification and sequencing.

It is important to point out that each sequencing method has its advantages and disadvantages, and understanding them is important in selecting the right method for your DNA damage detection requirements. RADAR-seq is the DNA damage detection method of choice to compare levels of different DNA lesions in a sample, to locate and quantitate rare DNA damage events (one in a million bases sequenced), or to quantitate the level of DNA damage produced by varying DNA damage conditions (as with increasing UV radiation in Basic Protocol 2).

As we have shown here, RADAR-seq can be utilized to assess the formation of DNA damage in an organism exposed to a DNA-damaging agent, and to assess the contribution of a DNA repair protein to in vivo DNA lesion repair. Here we use UV light as our DNA damage source to produce thymine dimers, but other DNA-damaging agents such as heat or oxidative agents can be used. Further, here we knock out RNaseH2 to assess its contribution to ribonucleotide excision repair in T. kodakarensis, but other DNA repair proteins can be assessed as long as the genetics are available and the repair protein is non-essential.

Critical Parameters

Reference genome

To perform RADAR-seq patch detection analysis, a reference genome is required to align RADAR-seq reads. It is important to perform library preparation, sequencing, and assembly for your strain(s) of interest as described in Basic Protocol 1 or to obtain a previously assembled reference genome (e.g., from NCBI). In RADAR-seq experiments in which a gene is being knocked out to assess its contribution to DNA damage and/or repair, it is important to confirm the knockout by sequencing before constructing a reference genome. For proper genome assembly, large-fragment PacBio libraries (>15 kb) are required. For genomes >12 Mb, a size selection step (e.g., Blue Pippin size selection) is routinely done at the end of PacBio library construction to ensure removal of any small insert DNAs from the library. Size selection is not required for smaller genomes, such as bacterial and archaeal genomes.

Genomic DNA quality

For RADAR-seq library construction and sequencing to be successful, it is imperative to start with enough high-quality, intact genomic DNA. Therefore, all genomic DNA undergoing RADAR-seq library construction should be quantitated using a Qubit (or Nanodrop) instrument and assessed for integrity using an Agilent TapeStation (or Agilent Bioanalyzer or agarose gel electrophoresis). It is important to start with at least 2 μg genomic DNA at the start of library construction, as DNA is lost during the construction process.

Many RADAR-seq experiments utilize genomic DNA extracted from an organism that was exposed to DNA-damaging agents. If exposure is too harsh, however, it can lead to highly fragmented genomic DNA, which results in difficult library construction or stalling of the PacBio sequencing polymerase during SMRT sequencing. It can thus be useful to perform a growth curve (if the organism allows) following DNA damage exposure to ensure that treatment does not lead to complete cellular apoptosis. One can proceed with library construction and sequencing of highly damaged DNA, but it may result in poor library yield and poor sequencing output.

Shearing of genomic DNA

PacBio library construction creates sequencing molecules with hairpin adapters to enable the sequencing polymerase to perform circular consensus sequencing (Eid et al., 2009), which allows the polymerase to sequence the same molecule many times. With each pass of the molecule, the sequencing results become more accurate. This includes a more accurate interpulse duration assessment for each base. RADAR-seq requires shearing of genomic DNA to 2-kb fragments to allow the PacBio sequencing polymerase to sequence each molecule with at least 3× coverage to ensure that calling of methylated bases is correct and, therefore, that detection of the patches of methylated bases is accurate. If...
3× coverage is not achieved for a particular molecule, the sequencing data are tossed from the analysis pipeline. The PacBio sequencing polymerase can achieve an average read length of 40 kb; therefore, a 2-kb fragment will have an average of ten passes (2 kb top strand + 2 kb bottom strand = 4 kb total). It is therefore important to confirm the size of the sheared genomic DNA (with an Agilent TapeStation, Agilent Bioanalyzer, or agarose gel electrophoresis) to ensure that the fragment size will enable at least three passes by the polymerase.

**PacBio library construction and BER enzymes**

PacBio library construction and RADAR-seq BER rely on many enzymatic reactions, each of which is critical for proper library construction. It is important that all enzymes be used within their shelf life to ensure activity. Inactivity of one enzyme in the library construction process can lead to low library yield or inaccurate results. For example, as discussed above, if Taq ligase has lost activity, ligation will be inefficient, leading to loss of fragments containing DNA lesions and a reduced number of detected DNA lesion sites.

**Sample handling**

It is important to avoid exposing samples to DNA-damaging conditions throughout library construction to avoid introduction of nicks or other DNA lesions. Therefore, we recommend minimizing freeze-thawing of samples and exposure to UV light and heat. It is best to perform the entire library construction portion of RADAR-seq succinctly to avoid formation of unwanted DNA nicks. Once the libraries have been constructed, they are safe to store at −20°C prior to PacBio sequencing.

**PacBio base modification detection versus RADAR-seq patch detection**

Because PacBio SMRT sequencing can distinguish 4mC and 6mA from canonical cytosine and adenine, it is routinely used to determine genomic 4mC and 6mA methylation motifs (Flusberg et al., 2010). Further, the PacBio SMRTLink Web portal provides a Base Modification tool that allows users to extract 4mC and 6mA methylation motifs from PacBio sequencing. It is important to point out that the standard Base Modification analysis tool on the PacBio SMRTLink Web portal is designed to detect methylation motifs within genomic DNA. These motifs are single base modifications that occur in a defined DNA sequence context. Because RADAR-seq produces a patch of methylated patches at DNA damage sites and not at defined sequence contexts, the standard Base Modification analysis cannot be used to extract the patches produced by RADAR-seq data. For this reason, the scripts provided here must be used to extract patches from the sequencing data.

**Troubleshooting**

For a list of common problems with the protocols, their causes, and potential solutions, please see Table 2.

**Understanding Results**

In our example of constructing a reference genome from MG1655 *E. coli* (see Basic Protocol 1), the integrity of the extracted genomic DNA was first checked using an Agilent TapeStation to ensure that it is high-molecular-weight and intact (Fig. 2A, lane 2). It was then sheared into 20-kb fragments using a Diagenode Megaruptor (Fig. 2A, lane 3). PacBio libraries were constructed and sequenced using a single SMRT Cell and PacBio Sequel instrument (Fig. 2B). The sequencing run produced 8,756,480,808 sequenced bases from the single SMRT Cell, with an average insert fragment length of 17,752 bases, which is in agreement with the 20-kb shearing protocol. The sequenced bases were used to construct a single circular contig with a length of 4,638,885 bases, equating to 1887.6-fold genome coverage. A single contig is expected for *E. coli*, which has a single circular chromosome, and the high genome coverage was achieved due to the small size of the *E. coli* genome. Assembly of larger genomes, especially eukaryotic genomes, will require multiple Sequel SMRT Cells to obtain enough sequencing data (see Strategic Planning and see Critical Parameters discussion of the reference genome).

In the RADAR-seq analysis of thymine dimers in *E. coli* exposed to varying levels of UV radiation (see Basic Protocol 2), the patch detection workflow produces a table with the genomic position, strand, and DNA sequence for all detected patches, where each patch represents one thymine dimer. From this table, one can sum the number of detected patches and map the location of each patch across the genome. For example, following the procedures described in Basic Protocol 2, we determined that the in vivo frequency of thymine dimers in *E. coli* genomic DNA without UV exposure was very low (~6 per million bases, Fig. 3C, Table 3), as expected. This confirms that thymine dimers were efficiently repaired by endogenous *E. coli*
| Problem                                                                 | Possible cause                                                                 | Solution                                                                 |
|------------------------------------------------------------------------|-------------------------------------------------------------------------------|-------------------------------------------------------------------------|
| Low yield from genomic DNA extraction                                  | Too much or too little tissue/culture input into genomic DNA extraction purification | Ensure input meets extraction kit requirements                          |
| Multiple contigs produced (or more than desired) during PacBio genome assembly | Poor-quality genomic DNA                                                      | Use freshly extracted, high-quality genomic DNA                          |
| DNA shearing produced smaller fragments than desired                    | Poor genomic DNA quality                                                      | Assess genomic DNA quality using TapeStation prior to shearing          |
| DNA shearing produced larger fragments than desired                     | Incorrect Covaris shearing settings                                           | Confirm settings or optimize to obtain correct fragment size             |
| Low yield after PacBio library prep                                     | DNA input too low                                                             | Re-quantitate genomic DNA; ensure integrity using Agilent TapeStation |
| Low yield after RADAR-seq BER                                           | Inactive DNA polymerase or ligase                                             | Ensure polymerase/ligase are not expired and are used in the correct ratio |
|                                                                        | Methyalted dNTPs or NAD\(^+\) excluded from reaction                        | Ensure dNTPs and NAD\(^+\) were added to reaction                      |
|                                                                        | Exonuclease contamination                                                     | Confirm cleanup was performed after exonuclease treatment and before BER |
| Low sequencing output                                                   | DNA input too low                                                             | Re-quantitate final library; if concentration in confirmed, add 1/3 more sample during sequencing |
|                                                                        | Contamination from library prep                                              | Perform another AMPure bead cleanup on final library                    |
|                                                                        | Sequencing polymerase is inactive/unbound                                     | Repeat primer/polymerase binding and sequencing                        |
| High damage in control experiments                                     | Low-integrity input DNA                                                       | Ensure integrity of input DNA using Agilent TapeStation                 |

(Continued)
Table 2 Troubleshooting Guide for RADAR-seq, continued

| Problem                        | Possible cause                                      | Solution                                               |
|--------------------------------|-----------------------------------------------------|--------------------------------------------------------|
| Omitted exonuclease step       | Confirm exonuclease step was performed prior to BER step (any nick present in sample will be translated to patch of modified bases) |
| Sequencing output did not map | Incorrect reference genome                          | Ensure correct reference genome                        |
|                                | Incorrect barcodes                                  | Ensure correct barcodes were chosen during sequencing setup |

Table 3 Number of Thymine Dimers Detected by RADAR-seq in *E. coli* Exposed to UV Radiation

| UV dosage (J/m²) | Total bases sequenced | Genome coverage | Thymine dimers detected per Mb | Thymine dimers detected per genome | Replicates |
|------------------|-----------------------|-----------------|-------------------------------|-----------------------------------|------------|
| 0                | 1,351,235,932         | 146 ± 6         | 6 ± 2                         | 55 ± 21                           | 3          |
| 10               | 1,423,858,251         | 153 ± 43        | 31 ± 9                        | 287 ± 81                          | 3          |
| 25               | 1,659,536,815         | 179 ± 35        | 45 ± 18                       | 415 ± 163                         | 4          |
| 50               | 610,458,437           | 66 ± 23         | 82 ± 20                       | 757 ± 189                         | 4          |
| 100              | 336,171,217           | 36 ± 8          | 113 ± 31                      | 1048 ± 286                        | 4          |
| 200              | 70,411,338            | 8 ± 1           | 232 ± 18                      | 2153 ± 166                        | 4          |

*All libraries were sequenced on a PacBio Sequel instrument. Individual sequencing runs were combined for data analysis.

*Mean ± standard deviation from sequencing replicates. Frequency per genome was determined as 2 × genome size × (frequency per Mb/1,000,000).

Repair pathways in the absence of UV exposure. With increasing dosages of UV radiation (10-200 J/m²²), we observed a linear increase in thymine dimer formation (Fig. 3B, Table 3). Importantly, under these experimental conditions, 10 J/m² was the lower limit of UV dosage required to produce at least a two-fold increase in thymine dimers compared to background unexposed DNA, providing the lower limit of thymine dimer detection by RADAR-seq, as cells exposed to 5 J/m² had background levels of thymine dimers (data not shown). Exposure at 200 J/m² resulted in a large increase in thymine dimers across the genome compared to unexposed DNA (232 thymine dimers per million bases or 2162 thymine dimers per *E. coli* genome; Table 3). This dosage damaged 0.3% of the 677,590 modifiable TT sites in the *E. coli* genome, corresponding to a frequency of one thymine dimer every 4360 bases. The location of each thymine dimer detected by RADAR-seq was plotted on a genome-wide scale using Circos software (Krzywinski et al., 2009) to give a circular plot that displays the distribution of thymine dimers on the outer strand (red) and inner strand (blue) (Fig. 3C, 3D). This analysis enables us to determine DNA damage hot spots.

Importantly, in addition to exogenous DNA-damaging agents such as UV radiation, endogenous mechanisms can facilitate formation or incorporation of DNA damage. To exemplify this, we utilized RADAR-seq to locate and quantitate ribonucleotides in WT versus ΔRNaseH2 *T. kodakarenensis* strains (see Basic Protocol 3; Heider, Burkhart, Santangelo, & Gardner, 2017). In the RADAR-seq workflow, RNaseH2 was used to create a nick at ribonucleotides in the DNA, followed by nick translation with methylated bases and DNA ligation (Fig. 4A). Following PacBio Sequencing and RADAR-seq patch detection, the number of ribonucleotides in both strains was quantitated (Table 4). ΔRNaseH2 *T. kodakarenensis* had a 10-fold increase in embedded ribonucleotides compared to wildtype (Fig. 4B; Table 4). The location of ribonucleotides was plotted on a genome-wide scale using Circos software (Krzywinski et al., 2009). Importantly, ribonucleotides were distributed equally across the genome, demonstrating that incorporation of ribonucleotides by the leading and lagging strand replicative polymerase is not sequence biased and that incorporation is random (Fig. 4C, 4D).
Table 4  Number of Ribonucleotides Detected by RADAR-seq in WT and ΔRNaseH2 T. kodakarensis

| Strain       | Total bases sequenced | Genome coveragea | Ribonucleotides detected per Mb b | Ribonucleotides detected per genome b | Replicates |
|--------------|-----------------------|-------------------|----------------------------------|---------------------------------------|------------|
| WT           | 753,412,275           | 180 ± 31          | 26 ± 7                           | 109 ± 27                              | 3          |
| ΔRNaseH2     | 850,213,785           | 204 ± 18          | 229 ± 35                         | 955 ± 146                             | 3          |

a All libraries were sequenced on a PacBio Sequel instrument. Individual sequencing runs were combined for data analysis.
b Mean ± standard deviation from sequencing replicates. Frequency per genome was determined as 2 × genome size × (frequency per Mb/1,000,000).

Table 5  RADAR-seq Experimental Planning

| Task                                      | Day 1         | Day 2                  | Day 3                  | Day 4                   | Day 5                   |
|-------------------------------------------|---------------|------------------------|------------------------|-------------------------|-------------------------|
| Organism growth                           | Morning       | Morning                | Morning                | Morning                 | Morning                 |
| Damage exposure (optional)                | Afternoon     | Afternoon              | Afternoon              | Afternoon               | Afternoon               |
| Genomic DNA extraction                    |               |                        |                        |                         |                         |
| Genomic DNA shearing                      |               |                        |                        |                         |                         |
| PacBio library construction (Part 1)      |               |                        |                        |                         |                         |
| PacBio library construction (Part 2)      |               |                        |                        |                         |                         |
| RADAR-seq base excision repair            |               |                        |                        |                         |                         |
| PacBio primer/polymerase binding          |               |                        |                        |                         |                         |
| PacBio sequencing                         |               |                        |                        |                         |                         |
| Patch detection analysis                  |               |                        |                        |                         |                         |

Time Considerations

Each protocol requires ~4 days from start to finish, including genomic DNA extraction through patch detection analysis. A breakdown of times for different steps includes: organism growth (overnight), genomic DNA extraction (1 hr), Qubit DNA quantitation (5 min/sample), quality assessment on the Agilent TapeStation (30 min), DNA shearing (15 min/sample). AMPure bead cleanup (30 min), library construction (3 hr), exonuclease digestion (30 min), BER (30-90 min depending on DNA lesion), primer/polymerase binding (2.5 hr), PacBio sequencing (>12.5 hr, depending on number of samples and movie length), and patch detection analysis (2 hr-2 days depending on sequencing output, genome size, and computational setup). Samples cannot be sheared simultaneously, so four samples will require 60 min. For additional notes on RADAR-seq computational optimization, see https://github.com/potapovneb/CP-RADAR-seq.

A typical schedule for performing RADAR-seq experiments is shown below. This will differ depending on the organism of interest and the number of RADAR-seq libraries that need to be constructed and sequenced. A Gantt chart is also provided to aid in experimental planning (Table 5).

Day 1
- Organism growth
- DNA damage exposure and repair experiments (optional)

Day 2, morning
- Genomic DNA extraction
- DNA quantitation and quality assessment
- DNA shearing
- DNA quantitation and quality assessment

Day 2, afternoon
- AMPure bead cleanup
- PacBio library preparation (DNA prep, DNA repair, end repair, DNA ligation (overnight)

Day 3, morning
- AMPure bead cleanup
- Exonuclease digestion
- AMPure bead cleanup and DNA quantitation
- BER (nicking DNA, incorporation modified nucleotide patch, sealing nick with Taq ligase)
- AMPure bead cleanup and DNA quantitation

Day 3, afternoon
- Primer/polymerase binding
- Instrument loading and sequencing (overnight)

Day 4
- PacBio sequencing continues
- Patch detection analysis

Acknowledgments
The authors would like to thank Brett Burkhart, Geraldy Liman, and Dr. Tom Santangelo for providing T. kodakarensis strains; Dr. Richard Morgan for helpful manuscript...
discussion; and the leadership of New England Biolabs for fostering a supportive research environment.

Author Contributions

Kelly M. Zatopek: Conceptualization, data curation, formal analysis, investigation, methodology, validation, visualization, writing (original draft, review and editing); Vladimir Potapov: Data curation, formal analysis, methodology, software, validation, visualization, writing (original draft, review and editing); Jennifer L. Ong: Conceptualization, methodology, supervision; Andrew F. Gardner: Conceptualization, methodology, supervision, writing (review and editing).

Conflict of Interest

The authors are employed and funded by New England Biolabs, a manufacturer and vendor of molecular biology reagents, including DNA repair enzymes. The authors state that this affiliation does not affect their impartiality, objectivity of data generation or interpretation, adherence to journal standards and policies, or availability of data.

Data Availability Statement

PacBio sequencing data generated from Basic Protocols 1-3 have been deposited in the Sequencing Read Archive (https://www.ncbi.nlm.nih.gov/sra) under accession number SRP373231.

Literature Cited

Bryan, D. S., Ransom, M., Adane, B., York, K., & Hesselberth, J. R. (2014). High resolution mapping of modified DNA nucleobases using excision repair enzymes. Genome Research, 24, 1534–1542. doi: 10.1101/gr.174052.114

Chatterjee, N., & Walker, G. C. (2017). Mechanisms of DNA damage, repair, and mutagenesis. Environmental and Molecular Mutagenesis, 58, 235–263. doi: 10.1002/em.22087

Chen, L., Liu, P., Evans, T. C. Jr., & Ettwiller, L. M. (2017). DNA damage is a pervasive cause of sequencing errors, directly confounding variant identification. Science, 355, 752–756. doi: 10.1126/science.aai8690

Claussen, A. R., Lujan, S. A., Burkholder, A. B., Orebaugh, C. D., Williams, J. S., Claussen, M. F., … Kunkel, T. A. (2015). Tracking replication enzymology in vivo by genome-wide mapping of ribonucleotide incorporation. Nature Structural & Molecular Biology, 22, 185–191. doi: 10.1038/nsmb.2957

Ding, J., Taylor, M. S., Jackson, A. P., & Reijns, M. A. (2015). Genome-wide mapping of embedded ribonucleotides and other noncanonical nucleotides using emRiboSeq and EndoSeq. Nature Protocols, 10, 1433–1444. doi: 10.1038/nprot.2015.099

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., … Turner, S. (2009). Real-time DNA sequencing from single polymerase molecules. Science, 323, 133–138. doi: 10.1126/science.1162986

Figueroa-Gonzalez, G., & Perez-Plasencia, C. (2017). Strategies for the evaluation of DNA damage and repair mechanisms in cancer. Oncology Letters, 13, 3982–3988. doi: 10.3892/ol.2017.6002

Flusberg, B. A., Webster, D. R., Lee, J. H., Travers, K. J., Olivares, E. C., Clark, T. A., … Turner, S. W. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. Nature Methods, 7, 461–465. doi: 10.1038/nmeth.1549

Friedberg, E. C. (2008). A brief history of the DNA repair field. Cell Research, 18, 3–7. doi: 10.1038/cr.2007.113

Gehring, A. M., Sanders, T. J., & Santangelo, T. J. (2017). Markerless gene editing in the hyperthermophilic archaeon Thermococcus kodakarensis. Bio-Protocol, 7, e2604. doi: 10.21769/BioProtoc.2604

Heider, M. R., Burkhart, B. W., Santangelo, T. J., & Gardner, A. F. (2017). Defining the RNaseH2 enzyme-initiated ribonucleotide excision repair pathway in archaea. Journal of Biological Chemistry, 292, 8835–8845. doi: 10.1074/jbc.M117.783472

Huang, R., & Zhou, P. K. (2021). DNA damage and repair mechanisms in cancer therapy. Nature Review Cancer, 21, 1071–1078. doi: 10.1038/s41591-021-00648-7

Jackson, S. P., & Bartek, J. (2009). The DNA-damage response in human biology and disease. Nature, 461, 1071–1074. doi: 10.1038/nature08467

Joyce, C. M. (1997). Choosing the right sugar: How polymerases select a nucleotide substrate. Proceedings of the National Academy of Sciences of the United States of America, 94, 1619–1622. doi: 10.1073/pnas.94.5.1619

Kellner, V., & Luke, B. (2020). Molecular and physiological consequences of faulty eukaryotic ribonucleotide excision repair. EMBO Journal, 39, e102309. doi: 10.15252/embj.2019102309

Keszthelyi, A., Daigaku, Y., Ptasinska, K., Miyabe, I., & Carr, A. M. (2015). Mapping ribonucleotides in genomic DNA and exploring replication dynamics by polymerase usage...
sequencing (Pu-seq). *Nature Protocols*, 10, 1786–1801. doi: 10.1038/nprot.2015.116

Kochiwa, H., Tomita, M., & Kanai, A. (2007). Evolution of ribonuclease H genes in prokaryotes to avoid inheritance of redundant genes. *BMC Evolutionary Biology*, 7, 128. doi: 10.1186/1471-2148-7-128

Koh, K. D., Balachander, S., Hesselberth, J. R., & Storici, F. (2015). Ribose-seq: Global mapping of ribonucleotides embedded in genomic DNA. *Nature Methods*, 12, 251–257. doi: 10.1038/nmeth.3259

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., … Marra, M. A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Research*, 19, 1639–1645. doi: 10.1101/gr.092759.109

Mao, P., Smerdon, M. J., Roberts, S. A., & Wyrick, J. J. (2016). Chromosomal landscape of UV damage formation and repair at single-nucleotide resolution. *Proceedings of the National Academy of Sciences of the United States of America*, 113, 9057–9062. doi: 10.1073/pnas.1606671113

McElhinny, S. A. N., Watts, B. E., Kumar, D., Watt, D. L., Lundstrom, E. B., Burgers, P. M., … Kunkel, T. A. (2010). Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 4949–4954. doi: 10.1073/pnas.0914857107

Minei, R., Hoshina, R., & Ogura, A. (2018). *De novo* assembly of middle-sized genome using MinION and Illumina sequencers. *BMC Genomics*, 19, 700. doi: 10.1186/s12864-018-5067-1

Mingard, C., Wu, J., McKeague, M., & Sturla, S. J. (2020). Next-generation DNA damage sequencing. *Chemical Society Reviews*, 49, 7354–7377. doi: 10.1039/D0CS00647E

Orebaugh, C. D., Lujan, S. A., Burkholler, A. B., Clausen, A. R., & Kunkel, T. A. (2018). Mapping ribonucleotides incorporated into DNA by hydrolytic end-sequencing. *Methods in Molecular Biology*, 1672, 329–345. doi: 10.1007/978-1-4939-7306-4_23

Ostling, O., & Johanson, K. J. (1984). Micro-electrophoretic study of radiation-induced DNA damages in individual mammalian cells. *Biochemical and Biophysical Research Communications*, 123, 291–298. doi: 10.1016/0006-291X(84)90411-X

Sloan, D. B., Broz, A. K., Sharbrough, J., & Wu, Z. (2018). Detecting rare mutations and DNA damage with sequencing-based methods. *Trends in Biotechnology*, 36, 729–740. doi: 10.1016/j.tibtech.2018.02.009

Wallace, S. S. (2014). Base excision repair: A critical player in many games. *DNA Repair*, 19, 14–26. doi: 10.1016/j.dnarep.2014.03.030

Zatopek, K. M., Potapov, V., Maduzia, L. L., Alpaslan, E., Chen, L., Evans, T. C. Jr., … Gardner, A. F. (2019). RADAR-seq: A RAre DAmage and Repair sequencing method for detecting DNA damage on a genome-wide scale. *DNA Repair*, 80, 36–44. doi: 10.1016/j.dnarep.2019.06.007

**Internet Resources**

https://github.com/potapovneb/CP-RADAR-seq; 10.5281/zenodo.6914938

RADAR-seq computational workflow in the GitHub repository.