Title

Characterization of R-Loop Structures Using Single-Molecule R-Loop Footprinting and Sequencing

Authors

Malig, Maika
Chedin, Frederic

Publication Date

2020

DOI

10.1007/978-1-0716-0680-3_15

Peer reviewed

HHS Public Access

Author manuscript

Methods Mol Biol. Author manuscript; available in PMC 2020 November 16.

Published in final edited form as:

Methods Mol Biol. 2020 ; 2161: 209–228. doi:10.1007/978-1-0716-0680-3_15.

Characterization of R-Loop Structures Using Single-Molecule R-Loop Footprinting and Sequencing

Maika Malig, **Frederic Chedin**

Department of Molecular and Cellular Biology, University of California, Davis, CA 95616

Abstract

R-loops are three-stranded structures that form during transcription when the nascent RNA hybridizes with the template DNA resulting in a DNA:RNA hybrid and a looped-out single-stranded DNA (ssDNA) strand. These structures are important for normal cellular processes and aberrant R-loop formation has been implicated in a number of pathological outcomes, including certain cancers and neurodegenerative diseases. Mapping R-loops has primarily been performed using DRIP (DNA:RNA immunoprecipitation) based methods that are dependent on the anti-DNA:RNA hybrid S9.6 antibody and short-read sequencing. While DRIP-based methods are robust and report R-loop formation genome-wide, they only do so at the population average level; interrogating R-loop formation at the single molecule level is not feasible with such approaches. Here we present single molecule R-loop footprinting (SMRF-seq), a method that relies on the chemical reactivity of the displaced ssDNA strand to non-denaturing sodium bisulfite and single molecule long-read sequencing as a readout, to characterize R-loops. SMRF-seq can be used independently of S9.6 to generate high resolution, strand-specific, maps of individual R-loops at ultra-deep coverage on kilobases-length DNA fragments.

Keywords

R-loop; DNA:RNA hybrid; Non-denaturing bisulfite conversion; SMRT sequencing; Transcription

1 Introduction

R-loops are three-stranded nucleic acid structures consisting of a DNA:RNA hybrid and a displaced single stranded DNA. R-loops are abundant non-B DNA structures in genomes, covering 5–10% of the genomic space in organisms ranging from yeasts, plants, and mammals [1–6]. While evidence exists to support that R-loops can form *in trans* [7–9], nuclear R-loops are predominantly understood to form co-transcriptionally from the re-invasion of the nascent transcript into the duplex DNA template. This is evidenced from the overwhelming genic distribution of R-loops, their co-directionality and correlation with transcription, and conclusions from in vitro transcription reactions [10]. Investigations into the possible roles of R-loops have suggested that these structures influence a range of cellular processes [11–13]. Under normal conditions, R-loops participate in the regulation of gene expression [2, 14, 15], class switch recombination in B cells [16,17], and efficient

transcription termination [18,19]. Deregulation of R-loop metabolism has also been linked to genome instability [20–22] ultimately contributing to pathological conditions such as certain cancers and neurological disorders [23, 24].

R-loops have been mapped genome-wide using DNA:RNAimmunoprecipitation (DRIP) [1], a method that relies on the S9.6 anti-DNA:RNA hybrid antibody [25] and high-throughput short-read sequencing. Since its initial introduction, DRIP has been widely adopted and several variants of the method have been introduced to improve its resolution and strand-specificity [2, 6, 15]. While robust, these methods nonetheless possess few limitations. A first complication is due to the significant residual affinity of the S9.6 antibody for double-stranded RNA (dsRNA) [26]. This residual affinity can result in inaccurate R-loop maps when RNA strands are directly sequenced unless additional steps are taken to remove contaminating dsRNAs [5,27]. Secondly, DRIP, like ChIP approaches, only provides a population average view of R-loop formation derived from a large cell population. The use of high-throughput short-read sequencing technologies as read-outs further precludes obtaining any information on individual R-loops including their lengths and individual start and stop positions. To overcome these limitations and provide an S9.6-independent method to profile individual R-loops at ultra-deep coverage, we developed Single-Molecule R-loop Footprinting and sequencing (SMRF-seq) [28]. SMRF-seq relies on non-denaturing bisulfite conversion [16] to catalyze the efficient conversion of unpaired cytosines (C) located on the displaced ssDNA strand of R-loops to uracils (U). After long-range locus-specific PCR amplification, molecules amplified from the tagged displaced strand will carry patches of cytosine to thymine (T) conversions flanked by unmodified DNA. These patches represent R-loop “footprints”. The use of Pacific Bioscience’s Single-Molecule Real Time sequencing (SMRT-seq) on long PCR amplicons permits the characterization of collections of individual footprints at high coverage. Overall, SMRF-seq permits R-loop mapping at high resolution, in a strand-specific manner, on kilobases-long single-molecule reads and at high coverage. SMRF-seq, together with its accompanying analysis package Gargamel, will fuel new waves of investigations into the formation and function of R-loop structures.

2 Material

Prepare all reagents and dilutions using molecular biology-grade nanopure water (ddH_2O).

2.1 General

1. Tris-EDTA (TE) buffer: 10 mM Tris–HCL pH 8.0 and 5 mM EDTA.
2. 200 proof Ethanol.
3. Elution Buffer (EB): 10 mM Tris–HCl pH 8.5.
4. 1.5 mL and 2 mL Microcentrifuge tubes.
5. 8-strip PCR tubes.
6. Bench top Microcentrifuge.
7. DynaMag-2 beads (Life Technologies, 12321D).

8. AMPure XP beads (Beckman Coulter). Set at room temperature before use.
9. Mini tube rotator.

2.2 Nucleic Acid Purification

1. 15 mL conical tubes.
2. Proteinase K. Dilute to 20 mg/mL working solution.
3. Sodium dodecyl sulfate (SDS). Dilute to 20% working solution.
4. 5PRIME Phase Lock Gel Light 2 mL tubes.
5. Phenol:Choloroform:Isoamyl alcohol (25:24:1).
6. 3 M Sodium Acetate (NaOAc) pH 5.2.
7. Wide bore 200 and 1000 μL barrier tips.

2.3 Non-denaturing Bisulfite Treatment

1. DNA Methylation-Lightning.
2. Rotisserie oven.

2.4 Site-Specific Amplification

1. Thermocycler (optional: gradient-enabled block).
2. Phusion Hot Start DNA Polymerase.
3. 5× Phusion HF Buffer.
4. 5× Phusion GC Buffer.
5. dNTPs: Equimolar amounts of ATP, CPT, GTP, TTP. Dilute to 10 mM working solution.
6. 5 M Betaine.

2.5 SMRT-Bell Library Preparation

1. SMRTbell Template Prep Kit 1.0 (Pacific Biosciences, 100-259-100).
2. AMPure PB 5 mL (Pacific Biosciences, 100-265-900).

2.6 Quality Check

1. Spectrophotometer.
2. 2100 Bioanalyzer.
3. GelRed (Phenix Research Products, RGB-4102).
4. 50 × TAE buffer: For a 1 L stock solution, dissolve 242 g Tris base in 600 mL ddH_2O. Add 57.1 mL glacial acetic acid and 100 mL 0.5 M EDTA pH 8. Adjust volume to 1 L with ddH_2O. Dilute stock to 1× working concentration.

5. Agarose gel: dissolve molecular biology grade agarose to a final 0.7% (weight vol.) concentration in 1× TAE by heating. Cool to ~65 °C. Add 1 × final concentration of GelRed. Cast gel and insert comb.

3 Methods

Overview.

SMRF-seq exploits the presence of long stretches of ssDNA on the displaced strands of three-stranded R-loop structures. The intrinsic ssDNA character of this strand can be revealed upon treatment with sodium bisulfite under non-denaturing conditions [16]. This will lead to the conversion of susceptible cytosines to uracils within the displaced strand of R-loops. This genetic tag can subsequently be read out after PCR amplification and library construction by long read, single-molecule sequencing (SMRT-seq) (Fig. 1). Downstream data analysis steps are performed using the available Gargamel pipeline [28] to seek and capture single-molecule R-loop footprints (SMRFs). True R-loop footprints are expected to be strand-specific and sensitive to pre-treatment with Ribonuclease H [2,16], an enzyme that specifically degrades RNA in the context of DNA:RNA hybrids [29]. By contrast, spontaneous DNA breathing or strand-separation events will lead to conversion on both DNA strands in a manner insensitive to Ribonuclease H activity. Minimal length thresholds enforced during analysis serve to reinforce the distinction between characteristically long R-loops and short DNA breathing events [30]. Note that SMRF-seq can be carried out with or without preliminary R-loop enrichment with the S9.6 antibody without distortion of footprints [28]. The protocol described here does not include any S9.6 enrichment. Detailed step-by-step instructions for performing DRIP have been recently published [31].

3.1 Cell Harvest

This section is optimized for the human Ntera-2 and HEK293 cell lines grown in culture. The following steps can be easily adapted to other cell lines. Note that SMRF-seq can be easily performed on R-loops generated upon in vitro transcription of plasmids carrying R-loop prone sequences [32, 33]. If interested in such application, carry out in vitro transcription as described [32] and proceed directly to non-denaturing bisulfite treatment (Subheading 3.3) using the products of in vitro transcription reactions as initial material.

1. Passage cells 16 h before harvesting.
2. After 16 h of growth, count cells and measure cell viability. Optimal cells count should be around 5–6 million cells per sample with >90% viable counts. Cells should be no more than 80–90% confluent at harvest.
3. Add 1.5 mL trypsin to dislodge cells and incubate for 2–5 min at 37 °C until cells come off the plate.
4. Add 5 mL culture medium and pipette mix to remove clumps.
5. Transfer to a new 15 mL conical tube and spin for 3–5 min at 1000 rpm (~200 RCF) to pellet cells.
6. Aspirate media carefully without disturbing cell pellet.

7. Add 10 mL 1 × DPBS and pipette mix to resuspend cells.
8. Re-spin at 200 RCF for 3–5 min to pellet cells.
9. Aspirate supernatant off without disturbing pellet.

3.2 Nucleic Acid Purification

1. Add 875 μL TE to the pelleted cells. Pipette mix ~10× to resuspend.

2. Transfer 200 μL of resuspended cells into a new 1.5 microcentrifuge tube. Up to four DNA extractions can be performed from the cell material collected.

3. Cell lysis: add the following volumes to each tube:

12.5 μL	20% SDS
10 μL	Proteinase K

4. Invert 4–5× gently until lysed. The solution should appear clear and viscous.

5. Incubate in a 37 °C water bath for 2 h to permit proteinase K digestion. Invert mix and do a brief spin down every hour to remove condensation that forms on the top of the tube.

6. Spin down 2 mL Phase Lock tube (PLT) for 1 min at 12,000 RCF.

7. Add 177.5 μL TE to each Proteinase K treated sample for a total volume of 400 μL.

8. Transfer (pouring or pipetting) Proteinase K treated cell lysate into PLT directly.

9. Add 400 μL Phenol:Choloroform:Isoamyl alcohol to lysate. Invert gently a few times (*see* Note 1).

10. Spin down at 12,000 RCF for 5 min at room temperature or at 4 °C if possible (*see* Note 2).

11. Pour top aqueous layer containing nucleic acids (~400 μL) from PLT into new 2 mL tube.

12. Add the following to the 400 μL sample to precipitate nucleic acid:

40 μL	3 M sodium acetate (1:10 by volume)
1.0 mL	100% ethanol (2.5× volume)

13. Invert gently until DNA precipitate (5–10 min) (*see* Note 2).

1. Wear appropriate PPE when handling Phenol:Choloroform: Isoamyl alcohol as it is considered a hazardous material. Dispense in a chemical hood. Dispose of waste in designated sealed container.
2. Keeping reagents and spins cold will help with nucleic acid precipitation. When precipitating nucleic acids, DNA should be transparent after ethanol addition. If nucleic acid looks “milky”, continue to invert gently and incubate for another 5–10 min. If the cloudiness does not go away, repeat the phenol organic extraction or repeat isolation. Do not vortex samples.

14. Spool DNA using wide bore 1000 μL barrier tips and transfer to a new 1.5 mL tube while taking care not to carry over residual supernatant. Otherwise, remove as much supernatant as possible.

15. Add 1.5 mL cold 70% ethanol. Incubate 5–10 min on ice.

16. Remove supernatant and repeat 70% ethanol wash. Do not spin between washes.

17. Remove excess ethanol and air dry while tubes are inverted. This may take ~1 h depending on the size of the pellet.

18. Add 50 μL TE and incubate on ice for 30 min to resuspend DNA. Do not pipette to mix during that time as R-loops are sensitive to mechanical stress.

19. Using a wide bore 200 μL tip, pipette mix 3×.

20. Measure nucleic acid concentration using a spectrophonometer (i.e., Nanodrop). Proceed to bisulfite treatment within 24 h, keeping DNA at −20 °C (*see* Note 3).

3.3 Non-denaturing Bisulfite Treatment (*See* Note 4)

1. Pipette 1.5 μg of DNA from previous isolation bring volume to 20 μL with PCR grade water.
2. Add 130 μL Lightning conversion reagent to 20 μL template. Invert mix ~3×.
3. Incubate at 37 °C in a rotisserie oven for 2 h, protected from light.
4. Add 600 μL M-binding buffer to spin column.
5. Load 150 μL bisulfite-treated sample to column with buffer. Gently invert mix a few times.
6. Spin samples at 10,000 RCF for 30 s. Discard flow-through.
7. Add 100 μL M-wash buffer to column. Spin for 30 s.
8. Add 200 μL L-Desulphonation buffer to column.
9. Incubate for 20 min at room temperature. Then spin for 30 s. Discard flow-through.
10. Add 200 μL M-wash buffer to column. Spin for 30 s.

---

3. To preserve R-loop structures, keep temperatures low. Handle with care, avoid vortexing and unnecessary pipetting. R-loops are sensitive to mechanical stress and can spontaneously fall apart, reducing yields as a result. It is particularly essential to avoid any nicking or fragmentation of the displaced ssDNA strand as this will prevent the subsequent PCR amplification of that strand and compromise the recovery of information derived from that strand.

4. Conditions specified in the methylation kit were modified to enable non-denaturing R-loop footprinting. These modifications include omitting the initial denaturation step to ensure that only intrinsically single-stranded regions are targeted, and lowering the temperature to 37 °C during bisulfite treatment. In addition, treatment time was lowered to 2 h. These modifications may lead to a slight reduction in the efficiency of conversion compared to that typically required for CpG methylation profiling. We encourage users to use denatured spike-in controls to directly measure the conversion efficiency. Based on our data [28], conversion efficiencies are routinely in the 80–90% range, which is sufficient to allow for R-loop footprinting. Conversion efficiency is also influenced by the amount of nucleic acids to be treated, with an optimal recommended amount of 200–500 ng of input DNA, up to a maximum of 2 μg. The amount of input DNA also determines the number of PCR reactions that can be carried out post-treatment. Each PCR reaction optimally uses 25–50 ng of bisulfite-converted DNA to ensure the recovery of a diverse set of R-loop footprints. The amount of input DNA recommended here to go into the bisulfite reaction (1.5 μg) ensures good conversion efficiency and a strong recovery yield that permits at least 20 downstream PCR reactions.

11. Repeat wash. Air dry for 1 min.
12. Place columns in new 1.5 mL tubes.
13. Add 12.5 μL EB to elute sample. Wait 2 min. Spin for 30 s to collect samples.
14. Quantify DNA using spectrophotometer. Recovery should be around 80%.

3.4 Site-Specific Amplification (*See* Note 5)

1. Assemble PCR reaction in 8-strip PCR tubes as follows (*see* Note 6):

Reagent	Vol. per reaction (μL)	Final concentration
ddH_2O	14.75	
5× Phusion HF/GC buffer	5.0	1×
10 mM dNTPs	0.5	200 **μ**M
Template (10 ng/μL)	2.5	25 ng
Forward + reverse primers (10 **μ**M each)	2.0	0.8 **μ**M
Phusion U DNA polymerase (2 U/ **μ**L)	0.25	0.02 U/**μ**L
Total	**25.0**	

For difficult templates, use reaction conditions below:

Reagent	Vol. per reaction (μL)	Final concentration
ddH_2O	9.75	
5× Phusion HF buffer	5.0	1×
10 mM dNTPs	0.5	200 **μ**M
Template (10 ng/**μ**L)	2.5	25 ng
Forward + reverse primers (10 **μ**M each)	2.0	0.8 **μM**
5 M Betaine	5.0	1×
Phusion U DNA polymerase (2 U/μL)	0.25	0.02 U/μL
Total	**25.0**	

2. Perform gradient PCR using program below: (*see* Note 7).

---

5. Native primers are designed to hybridize to dsDNA regions flanking the candidate R-loops under study to ensure that no selection or distortion of R-loop patterns is introduced (Fig. 7). This is distinct from past protocols [16], where. converted primers that hybridize directly to the C to T converted strand were used to increase the recovery of R-loop structures. Typical amplicon lengths range from 2 to 5 kb, which permits efficient PacBio sequencing. For long-range amplicons (>1 kb), PCR optimization may be necessary especially for high GC content or repetitive regions.

6. Due to the high-throughput capacity of current PacBio sequencing instruments (4–500,000 reads on PacBio Sequel), it is recommended to pool multiple samples in one sequencing reaction. When pooling amplicons, we recommend that amplicons be in the same size range and suggest pooling non-overlapping genomic regions that can be uniquely mapped. In this way, barcoding is not necessary as mapping can easily be performed against the reference amplicons. In case users interested in mapping R-loops observed the same amplicon under varied conditions, barcoding is required. Barcoding is not covered in this protocol and users are referred to available technical information from PacBio (www.pacb.com).

7. When optimizing primers, use DNA from cell line to be used. In general, use HF buffer. For high GC content regions, use GC buffer. For difficult templates, use HF or GC buffer with betaine as a PCR additive. For gradient PCR, set a range of ±7 °C from the expected primer annealing temperatures to determine the optimum in terms of efficiency and specificity. It is preferable to perform the least possible number of cycles to decrease PCR duplicates and increase molecular diversity.

Stage	Temp (°C)	Time	Cycles
1. Denaturation	98	30 s	1
2. Amplification Denature	98	10 s	25–35
Annealing	50–72	30 s	
Extension	72	2.5 min	
3. Final extension	72	5 min	1
4. Hold	4	∞	

3. Check PCR products via agarose gel electrophoresis. The goal is to obtain a single PCR band of the correct size (Fig. 2). If a single band cannot be achieved, we advise the redesign and optimization of primers. In the event that contaminating bands cannot be completely reduced, then the dominant correct band can be excised from gels post PCR and the DNA purified subsequently.

4. Amplification of converted template. Use optimized PCR conditions as determined above. Double total reaction volume to 50 to increase recovery.

5. Purify products using AMPure Magbeads (*see* Note 8).

6. Transfer the 50 μL PCR product from the previous step into a new 1.5 mL tube.

7. Add 50 μL AMPure XP beads (1×) to the template.

8. Gently tap to mix. Briefly spin down.

9. Place samples in mini tube rotator and incubate for 10 min.

10. Briefly spin down then place tubes on DynaMag-2 magnetic rack. Wait ~1 min to allow beads migrate to the magnet and the supernatant to clear.

11. Remove supernatant carefully without pulling any beads.

12. Add 1.5 mL 70% ethanol. Cap tubes and invert a few times while on the magnetic rack. Remove supernatant.

13. Repeat ethanol wash.

14. Air dry for ~1 min. Do not let the bead dry excessively, which is evident when the beads begin to crack.

15. Resuspend in 30 μL EB, remove the tubes from the magnetic rack and gently mix by tapping or flicking. Incubate with rotation for 5 min.

16. Place tubes back on the DynaMag-2 rack and wait ~1 min. Recover DNA in supernatant and place in new 1.5 mL tube.

17. Check sample concentration using a spectrophotometer.

8. Avoid unnecessary pipetting to prevent damaging or fragmenting PCR products, especially for longer amplicons. Make sure that AMPure beads are at room temperature to avoid sample loss. Use freshly made 70% ethanol when performing bead washes.

18. Calculate number of molecules per amplicon:

$$\text{mole dsDNA(mol)} = \left(\frac{\text{mass of dsDNA(g)}}{\text{length of dsDNA(bp)} \times 650\left(\frac{\text{g}}{\text{mol.bp}}\right)} \right),$$

where the average weight of a DNA base-pair is 650 g/mol.

19. If pooling non-overlapping amplicons, pool equimolar amounts of each amplicon together with a total final mass no less than 500 ng for library preparation. Suggested starting for library preparation input is 500–700 ng.

3.5 SMRT-Bell Library Preparation (*See* Note 9)

1. Concentrate pooled amplicons (*see* Note 10).

2. Add 0.8× AMPure PB magnetic beads to pooled amplicons (i.e., if starting with 100 μL pooled sample, use 80 μL AMPure PB beads).

3. Perform ethanol purification as described in previous section for AMPure bead clean-up, **steps 8**–**16**.

4. Elute in 22 μL EB. Proceed to end repair step or store in −20 °C until ready to proceed to end-repair.

5. Assemble end-repair reaction on ice as follows:

Reagent	Volume (μL)	Final Conc.
Concentrated pooled amplicons	22.0	
10× template prep buffer	3.0	1×
10 mM ATP high	3.0	l mM
10 mM dNTP	1.2	0.4 mM
20× end repair mix	l.5	1×
Total volume	**30.0**	

6. Mix samples by gentle tapping or flicking the tube and briefly spin down.

7. Incubate samples at 25°C for 15 min in thermocycler, and then hold reaction at 4 °C.

9. The SMRTbell template library preparation presented here is for version 1 kits and adapted from PacBio's "Procedure & Checklist—2 kb Template Preparation and Sequencing" (PN 001-143-835-08). Estimated recovery after library prep is 20–30%. If preparing the library for the first time, we suggest starting with ~1 μg template and adjust adapter concentration (*see* Note 11). Newer version kits have different steps and may use a different set of primers for sequencing. Make sure to notify your PacBio sequencing provider which version of template prep kit you are using. For amplicons less than 5 kb in length, 10-h movie times should suffice to get at least 3 minimum DNA polymerase pass. Note that longer movie times can yield higher quality sequencing reads.

10. Optimize amplicons to avoid wide ranges in lengths when pooling amplicons for sequencing. In general, smaller fragments are more efficiently sequenced, leading to under-representation of longer fragments. Small fragment contaminations, such as from primer dimers, will soak up sequencable reads. Performing an additional AMPure clean-up, reducing the ratio of AMPure beads to DNA to 0.6–0.8× can help remove smaller fragments. See a recent protocol for more detailed instructions on how to perform AMPure cleanups [31]. Be sure to use AMPure BP beads during SMRTbell library preparation. Using generic AMPure beads may result in sequencing failure.

8. Purify end-repaired template using 0.8× AMPure PB beads (24 μL) and elute in 32 μL EB. This is a good stopping point. End-repaired samples can be stored at −20 °C for a few days. If continuing, proceed directly to adaptor ligation through the end of library preparation.

9. For blunt end SMRTbell adaptor ligation assemble reaction on ice as follows to avoid adapter-adapter ligation:

Reagent	Volume (μL)	Final Conc.
Repaired template	32.0	
20 μM blunt adapter	1.0	0.5 μM
Gently mix before next step		
10× template prep buffer	4.0	1×
1 mM ATP low	2.0	0.05 mM
Gently mix before next step		
Ligase (30 U/μL)	1.0	0.75 U/μL
Total volume	**40.0**	

10. Gently flick or tap to mix and spin down briefly.

11. Incubate for 1 h at 25°C in a thermocycler. Hold at 4 °C for at least 1 min, up to overnight (*see* Note 11).

12. Inactivate ligase by incubating at 65 °C for 10 min in a thermocycler. Hold at 4 °C for at least 1 min. Proceed directly to exonuclease treatment.

13. Exonuclease treatment. This will remove unligated and damaged template.

Set reaction on ice as follows:

Reagent	Volume (μL)
Ligated template	40.0
Exo III (100 U/**μ**L)	0.5
Exo VII (100 U/**μ**L)	0.5
Total volume	**41.0**

14. Gently flick or tap to mix reaction then briefly spin down.

15. Incubate for 1 h at 37 °C then hold at 4 °C in a thermocycler for at least 1 min. Proceed directly to clean-up step.

16. Library Clean-up, first purification. Perform purification using 0.8× AMPure PB magnetic beads (32.8 μL).

11. In the original PacBio protocol, ligation incubation can be extended to overnight. Extended incubation and increased input DNA may result in increased chimeric ligation products (double-insert templates), which can be checked on BioAnalyzer trace. If this happens, adjust adapter concentration to 30× (up to 50×) molar excess.

17. Elute in 50 μL EB then proceed directly to second purification.

18. Second purification, perform the last purification using 0.8× AMPure PB magnetic beads (40 μL).

19. Elute in 12 μL EB. This will be the library to be sent for sequencing.

20. Check concentration on a spectrophotometer. Check with your PacBio sequencing provider the requirements for library submission. If the concentration is too low, repeat library preparation from pooled amplicons.

21. Proceed to quality check. Use 1 μL of library to run on a BioAnalyzer High-Sensitivity chip (*see* Fig. 3 for example).

3.6 SMRF-Seq Analysis Using Gargamel Pipeline (*See* Note 12)

1. Download SMRT Link package from https://www.pacb.com/support/software-downloads/ (*see* Note 13) and instructions (https://www.pacb.com/wp-content/uploads/SMRT_Link_Installation_v600.pdf). CCS generation will be performed on the command line, thus GUI interface is not necessary. Perform all system checks to be able to call CCS tool.

2. Load smrtlink after opening a terminal for Circular Consensus Sequence (CCS) generation.

```
module load smrtlink/6.0
```

3. Run CCS tool. Depending on file size, this may take several hours to several days depending on compute capacity and file size. Input file will be under Primary Analysis folder annotated as “subreads.bam”. Default parameters are --minPasses = 3 and --minPredictedAccuracy = 0.9.

```
ccs <input.subreads.bam> <output.fq>
```

4. Removing PCR duplicates using dedupe2.sh from BBMap package v37.90 (https://sourceforge.net/projects/bbmap/). Thresholds had been optimized for this application specifically.

```
dedupe2.sh in=ccs.fq out=deduplicated.fq e=30 mid=98 k=31 nam=4
```

5. Analyze data using Gargamel pipeline. This is an open source software designed to call R-loop peaks and visualize patterns of C to T conversions. General usage for this

---

12. Further description of this analysis can be found here [28] and the GitHub page (listed in Subheading 3.6, **step 1**). This analysis pipeline can be computationally intensive depending on the size of data. We suggest using a high-performance computer cluster to submit jobs. It is highly suggested to use terminal multiplexer such as Screen or TMUX before starting a job. A newer version of SMRTLink v7 is available for download; this version should be compatible with previous versions. Software downloads can be requested from this link: http://www.pacb.com/support/software-downloads/

13. Thresholds are derived from [28]. The pipeline only considers reads if they are 95% of the total expected amplicon length. This may be too stringent for some purposes and can be reduced. 10 bp buffer flanking the start and end of amplicons are used to improve mapping as indicated by −x and −y. For the -l option, use descriptive alphanumerical values like ‘Experiment! to clearly annotate the outputs of successive data analyses.

pipeline is available at: https://github.com/srhartono/footLoop/blob/master/README.md

6. Download Gargamel pipeline:

git clonehttps://github.com/srhartono/footLoop

Required packages for this pipeline are listed in the README.md document.

7. Map and assign reads using footLoop (Bismark) (*see* Note 13).

This will require an index file and the reference genome to be mapped to (*see* Fig. 4 for example).

```
footLoop.pl –r <ccs.fq> –n <output_dir> –l <label> –i
<index.bed> –g <ref.genome.fa> –x –10 –y 10 –L 95p
```

8. Call C to T conversion tracks (footPeak) (*see* Note 14).

```
footPeak.pl –n <footLoop_output_dir> –o <output_dir> –w 20 –t 0.55–l 100
```

9. Generate peak clusters (footClust).

```
footClust.pl –n <footPeak_output_dir>
```

10. Visualize tracks by generating png files (footPeak_graph) (*see* Note 15).

```
footPeak_graph.pl –n <footPeak_output_dir> –r 1
```

(*See* Fig. 5 for an example)

11. Generate Genome browser tracks (footPeakGTF).

```
footPeak_GTF.pl –n <footPeak_output_dir>
```

This script will generate GTF files that can be directly uploaded to USCS Genome Browser. Example output is shown in Fig. 6.

Acknowledgements

We thank Chedin lab members for useful discussions, Dr. Lionel A. Sanz for constructive comments on the manuscript, and Dr. Stella R. Hartono for developing the Gargamel analysis pipeline. This work was funded by the National Institutes of Health (Grant R01 GM120607 to F.C.) and was supported, in part, by National Science

14. Peak calling is further described [28] and the GitHub page (listed in Subheading 3.6, **step 1**). The default threshold calls for at least 55% C to T conversion per window composed of 20 cytosines. A minimum 100 bp length was imposed. These thresholds can be varied by users. A key indicator that the threshold is working is that conversion patterns should be strand-specific for R-loops.

15. This script can create PNG and PDF files by using −r 1 and −R 1 options respectively. Option 1 creates files for the relevant data each parsed in its own appropriate directory (i.e., ‘peak’ and ‘no peak’ reads, ‘template’ and ‘non-template’ strand reads). Four relevant directories are generated. The ‘PEAK’ directory contains files for peak-containing reads from the non-template strand (i.e., converted looped-out DNA strand). The ‘NOPK’ directory contains no-peak reads from the non-template strand. The ‘PEAK_TEMP’ and ‘NOPK_- TEMP’ contain the template strand information. File names will contain “CH” and “GH” annotations indicating positive and negative strand, respectively.

Foundation Graduate Research Fellowship (Grant 1650042 to M.M.) and National Institute of General Medical Sciences Biomolecular Technology Predoctoral T32 Training Program (Grant T32-GM008799 to M.M.).

References

1. Ginno PA, Lott PL, Christensen HC, Korf I, Chedin F (2012) R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol Cell 45(6):814–825. 10.1016/j.molcel.2012.01.017 [PubMed: 22387027]
2. Sanz LA, Hartono SR, Lim YW, Steyaert S, Rajpurkar A, Ginno Pa, Xu X, Chedin F (2016) Prevalent, dynamic, and conserved R-loop structures associate with specific epigenomic signatures in mammals. Mol Cell 63 :167–178. 10.1016/j.molcel.2016.05.032 [PubMed: 27373332]
3. Wahba L, Costantino L, Tan FJ, Zimmer A, Koshland D (2016) S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. Genes Dev 30 :1327–1338. 10.1101/gad.280834.116 [PubMed: 27298336]
4. El Hage A, Webb S, Kerr A, Tollervey D (2014) Genome-wide distribution of RNA-DNA hybrids identifies RNase H targets in tRNA genes, retrotransposons and mitochondria. PLoS Genet 10(10):e1004716. 10.1371/journal.pgen.1004716
5. Hartono SR, Malapert A, Legros P, Bernard P, Chedin F, Vanoosthuyse V (2018) The affinity of the S9.6 antibody for double-stranded RNAs impacts the accurate mapping of R-loops in fission yeast. J Mol Biol 430 (3):272–284. 10.1016/j.jmb.2017.12.016 [PubMed: 29289567]
6. Xu W, Xu H, Li K, Fan Y, Liu Y, Yang X, Sun Q (2017) The R-loop is a common chromatin feature of the Arabidopsis genome. Nat Plants 3(9):704–714. 10.1038/s41477-017-0004-x [PubMed: 28848233]
7. Zaitsev EN, Kowalczykowski SC (2000) A novel pairing process promoted by Escherichia coli RecA protein: inverse DNA and RNA strand exchange. Genes Dev 14(6):740–749 [PubMed: 10733533]
8. Wahba L, Gore SK, Koshland D (2013) The homologous recombination machinery modulates the formation of RNA-DNA hybrids and associated chromosome instability, elife 2: e00505. 10.7554/eLife.00505
9. Kasahara M, Clikeman JA, Bates DB, Kogoma T (2000) RecA protein-dependent R-loop formation in vitro. Genes Dev 14(3):360–365 [PubMed: 10673507]
10. Chedin F (2016) Nascent connections: R-loops and chromatin patterning. Trends Genet 32(12):828–838. 10.1016/j.tig.2016.10.002 [PubMed: 27793359]
11. Santos-Pereira JM, Aguilera A (2015) Rloops: new modulators of genome dynamics and function. Nat Rev Genet 16(10):583–597. 10.1038/nrg3961 [PubMed: 26370899]
12. Costantino L, Koshland D (2015) The yin and Yang of R-loop biology. Curr Opin Cell Biol 34:39–45. 10.1016/j.ceb.2015.04.008 [PubMed: 25938907]
13. Crossley MP, Bocek M, Cimprich KA (2019) R-loops as cellular regulators and genomic threats. Mol Cell 73(3):398–411. 10.1016/j.molcel.2019.01.024 [PubMed: 30735654]
14. Chen L, Chen JY, Zhang X, Gu Y, Xiao R, Shao C, Tang P, Qian H, Luo D, Li H, Zhou Y, Zhang DE, Fu XD (2017) R-ChIP using inactive RNase H reveals dynamic coupling of R-loops with transcriptional pausing at gene promoters. Mol Cell 68(4):745–757. e745. 10.1016/j.molcel.2017.10.008
15. Chen PB, Chen HV, Acharya D, Rando OJ, Fazzio TG (2015) Rloops regulate promoter-proximal chromatin architecture and cellular differentiation. Nat Struct Mol Biol 22 :999–1007. 10.1038/nsmb.3122 [PubMed: 26551076]
16. Yu K, Chedin F, Hsieh CL, Wilson TE, Lieber MR (2003) R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. Nat Immunol 4(5):442–451. 10.1038/ni919 [PubMed: 12679812]
17. Wiedemann EM, Peycheva M, Pavri R (2016) DNA replication origins in immunoglobulin switch regions regulate class switch recombination in an R-loop-dependent manner. Cell Rep 17(11):2927–2942. 10.1016/j.celrep.2016.11.041 [PubMed: 27974207]

18. Skourti-Stathaki K, Proudfoot NJ, Gromak N (2011) Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. Mol Cell 42(6):794–805. 10.1016/j.molcel.2011.04.026 [PubMed: 21700224]
19. Proudfoot NJ (2016) Transcriptional termination in mammals: stopping the RNA polymerase II juggernaut. Science 352(6291): aad9926. 10.1126/science.aad9926
20. Stork CT, Bocek M, Crossley MP, Sollier J, Sanz LA, Chedin F, Swigut T, Cimprich KA (2016) Co-transcriptional R-loops are the main cause of estrogen-induced DNA damage. Elife 5:e17548. 10.7554/eLife.17548
21. Sollier J, Cimprich KA (2015) Breaking bad: R-loops and genome integrity. Trends Cell Biol 25(9):514–522. 10.1016/j.tcb.2015.05.003 [PubMed: 26045257]
22. Aguilera A, Garcia-Muse T (2012) R loops: from transcription byproducts to threats to genome stability. Mol Cell 46(2):115–124. 10.1016/j.molcel.2012.04.009 [PubMed: 22541554]
23. Richard P, Manley JL (2017) Rloops and links to human disease. J Mol Biol 429 (21):3168–3180. 10.1016/j.jmb.2016.08.031 [PubMed: 27600412]
24. Groh M, Gromak N (2014) Out of balance: R-loops in human disease. PLoS Genet 10(9): e1004630. 10.1371/journal.pgen.1004630
25. Boguslawski SJ, Smith DE, Michalak MA, Mickelson IKE, Yehle CO, Patterson WL, Carrico RJ (1986) Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. J Immunol Methods 89(1):123–130 [PubMed: 2422282]
26. Phillips DD, Garboczi DN, Singh K, Hu Z, Leppla SH, Leysath CE (2013) The sub-nanomolar binding of DNA-RNA hybrids by the single-chain Fv fragment of antibody S9.6. J Mol Recognit 26(8):376–381. 10.1002/jmr.2284 [PubMed: 23784994]
27. Vanoosthuyse V (2018) Strengths and weaknesses of the current strategies to map and characterize R-loops. Noncoding RNA 4(2): E9. 10.3390/ncrna4020009
28. Malig M, Hartono SR, Giafaglione JM, Sanz LA, Chedin F (2019) High-Throughput Single-Molecule R-loop Footprinting Reveals Principles of R-loop Formation. bioR-xiv:640094 10.1101/640094
29. Cerritelli SM, Crouch RJ (2009) Ribonuclease H: the enzymes in eukaryotes. FEBS J 276 (6):1494–1505 [PubMed: 19228196]
30. Kouzine F, Wojtowicz D, Baranello L, Yamane A, Nelson S, Resch W, Kieffer-Kwon KR, Benham CJ, Casellas R, Przytycka TM, Levens D (2017) Permanganate/S1 nuclease Footprinting reveals non-B DNA structures with regulatory potential across a mammalian genome. Cell Syst 4(3):344–356.e347. 10.1016/jxels.2017.01.013
31. Sanz LA, Chedin F (2019) High-resolution, strand-specific R-loop mapping via S9.6-based DNA-RNA immunoprecipitation and high- throughput sequencing. Nat Protoc 14 (6):1734–1755. 10.1038/s41596-019-0159-1 [PubMed: 31053798]
32. Stolz R, Sulthana S, Hartono SR, Malig M, Benham CJ, Chedin F (2019) Interplay between DNA sequence and negative superhelicity drives R-loop structures. Proc Natl Acad Sci U S A 116(13):6260–6269. 10.1073/pnas.1819476116 [PubMed: 30850542]
33. Carrasco-Salas Y, Malapert A, Sulthana S, Molcrette B, Chazot-Franguiadakis L, Bernard P, Chedin F, Faivre-Moskalenko C, Vanoosthuyse V (2019) The extruded non-template strand determines the architecture of R-loops. Nucleic Acids Res 47 (13):6783–6795. 10.1093/nar/gkz341 [PubMed: 31066439]

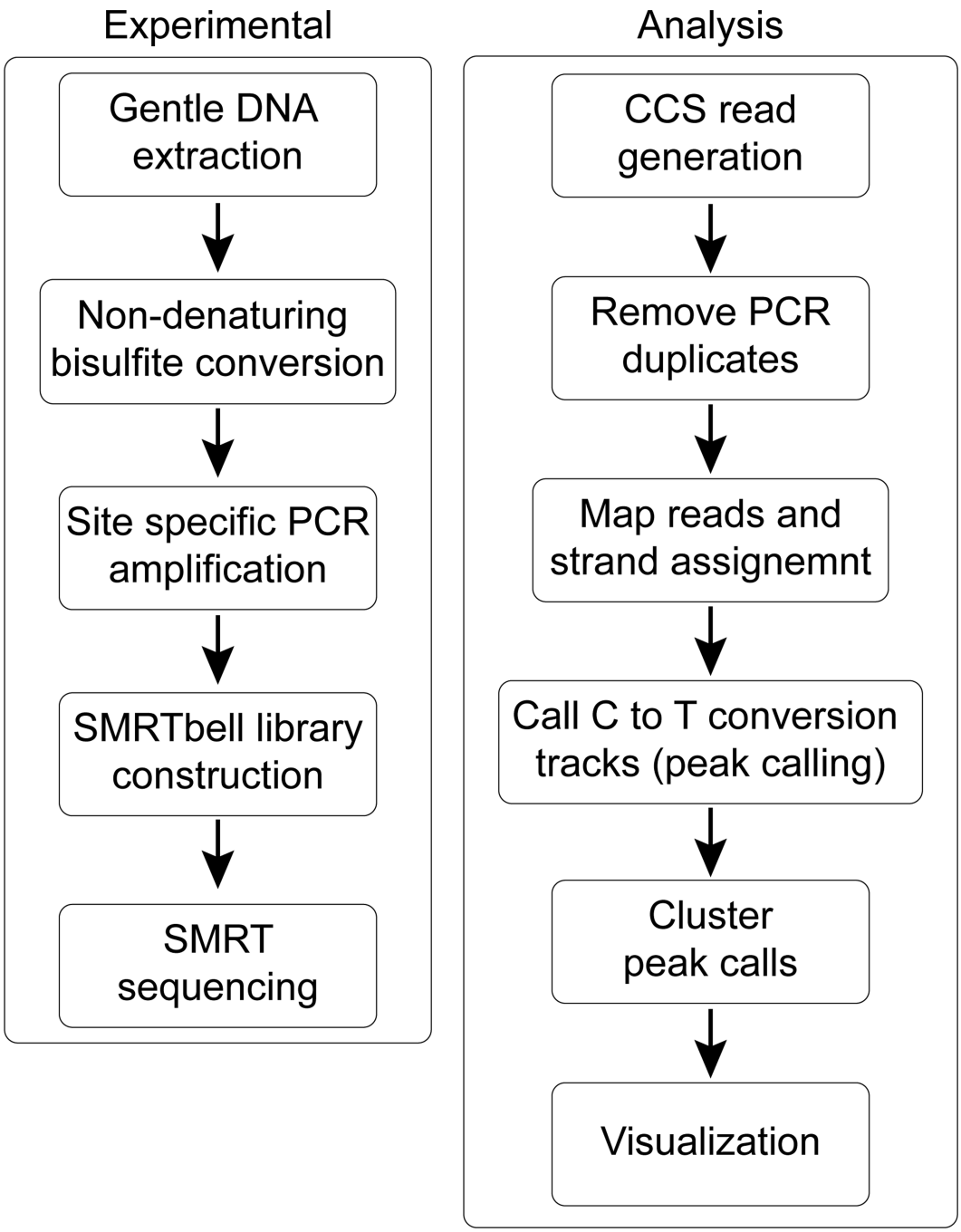

Fig. 1.
Overall workflow for both experimental procedures and computational analysis

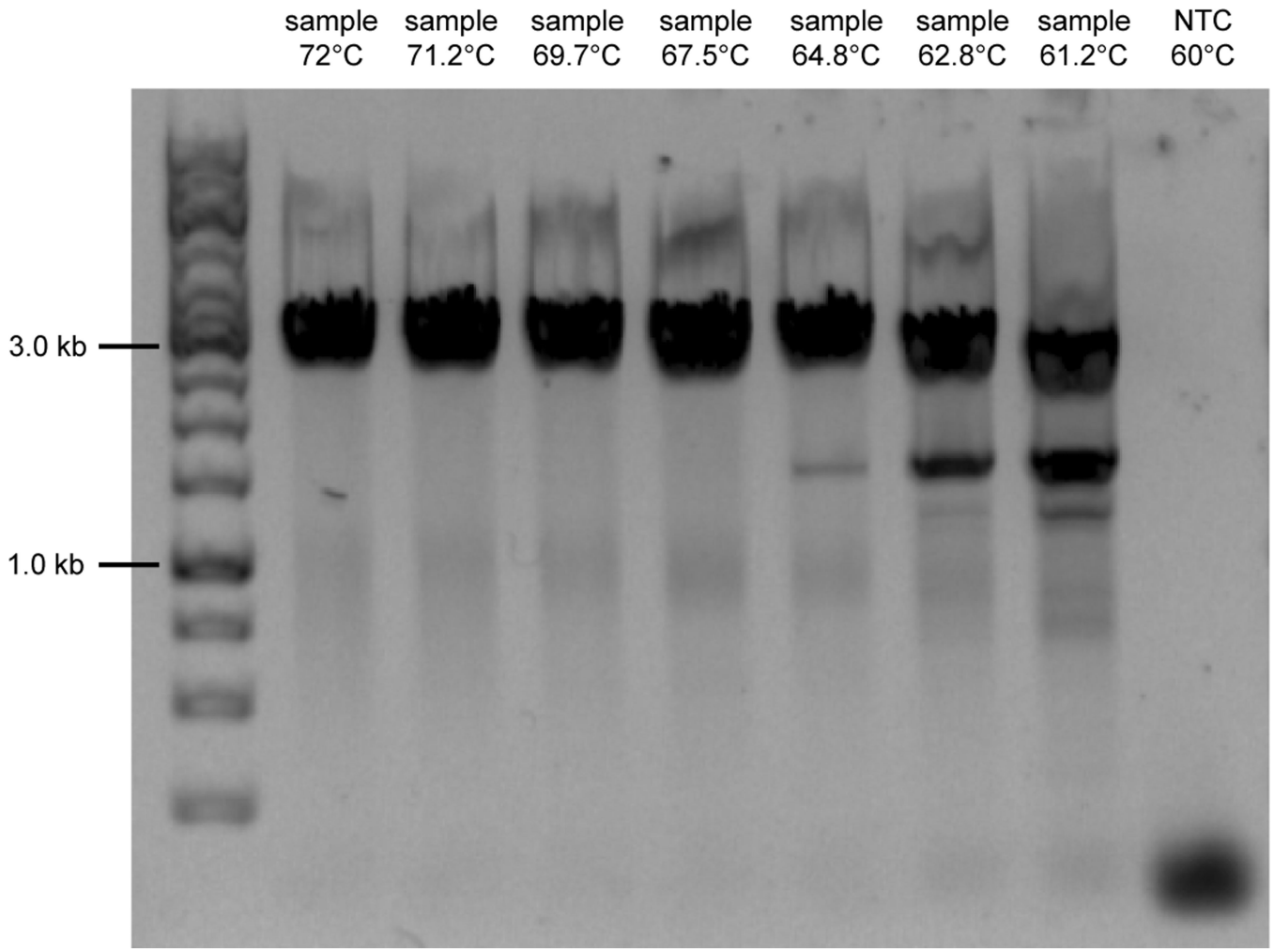

Fig. 2.
Example gradient PCR with 2990 bp product. Agarose gel stained with 1× GelRed. Run for 20 min at 150 V

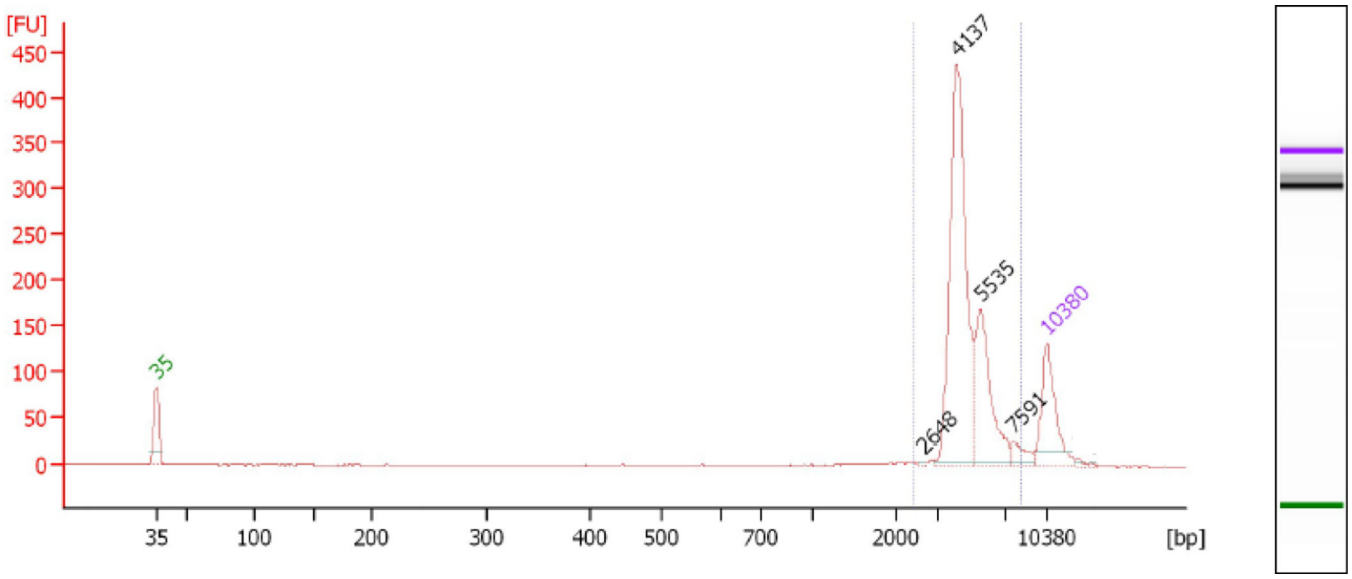

Fig. 3.

Example BioAnalyzer trace of a SMRTbell library. Pooled amplicons range from 2938 to 4379 bp

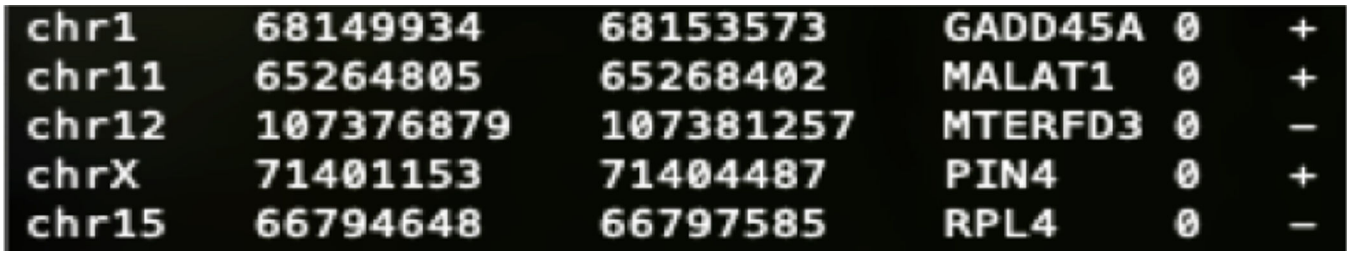

Fig. 4.

Example index file. Columns are as follows: chr, start, end, gene, 0, strand. Tab delimited and no header format

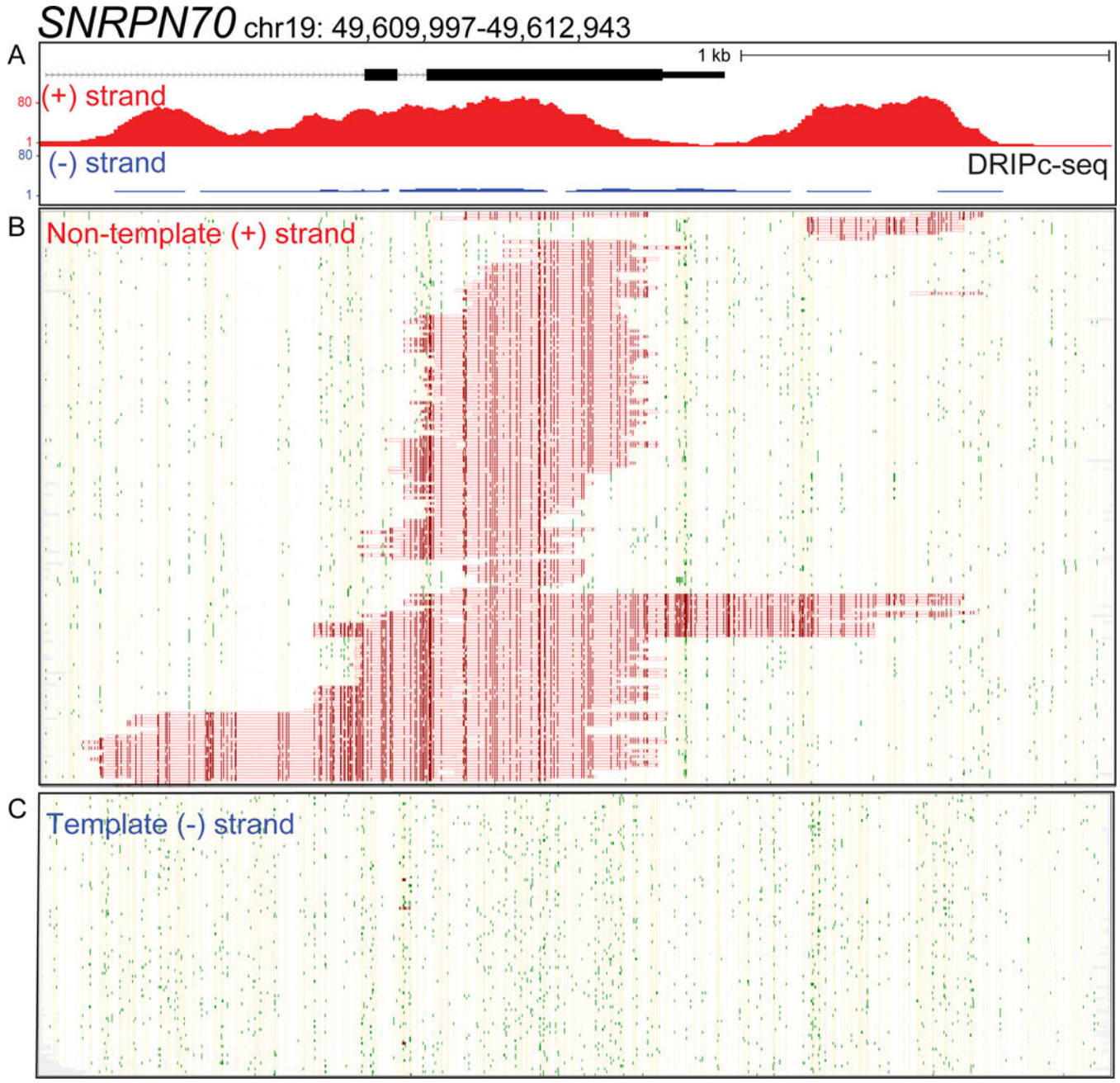

Fig. 5.

Example footprinted region with PNG output after footPeak_graph.pl run. (**a**) *SNRPN70* footprinted region (hg19) and DRIPc-seq data with (+) and (−) strands shown in red and blue, respectively. (**b**) Example output showing random 200 peaks reads on the non-template strand. (**c**) Example output showing random 100 reads for template strand. For panels (**b**) and (**c**), each horizontal line corresponds to one read. Red tick lines indicate converted Cs within a called peak, yellow indicates non-converted Cs, green indicates converted Cs not called a peak; gray is missing/ambiguous sequence

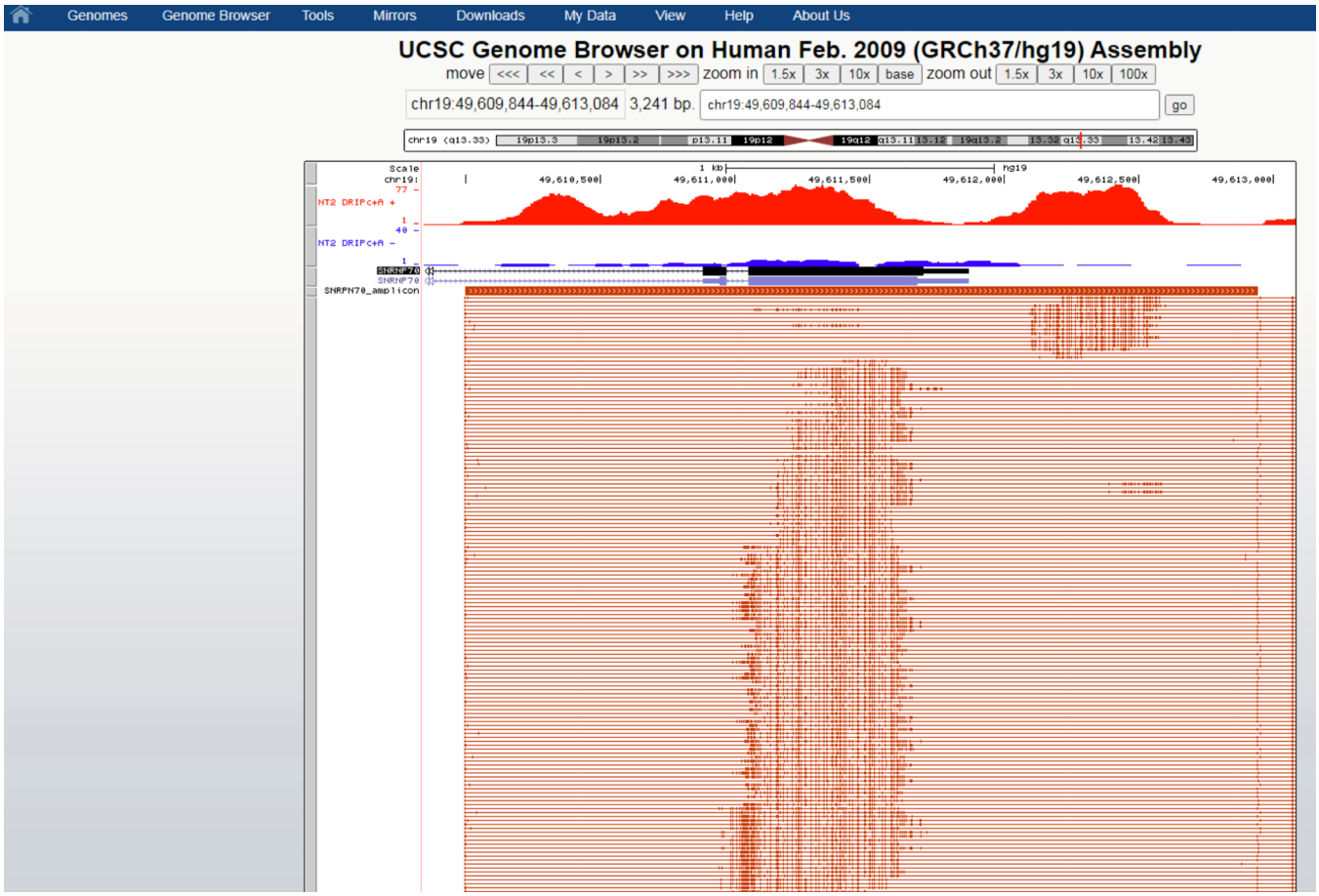

Fig. 6.

Genome Browser snapshot of same ***SNRPN70*** terminal region in previous figure. Converted cytosines within called peaks are annotated as red tick marks on the positive strand (blue tick marks for footprints called in the negative strand). Each horizontal line is a single molecule read mapping to the region

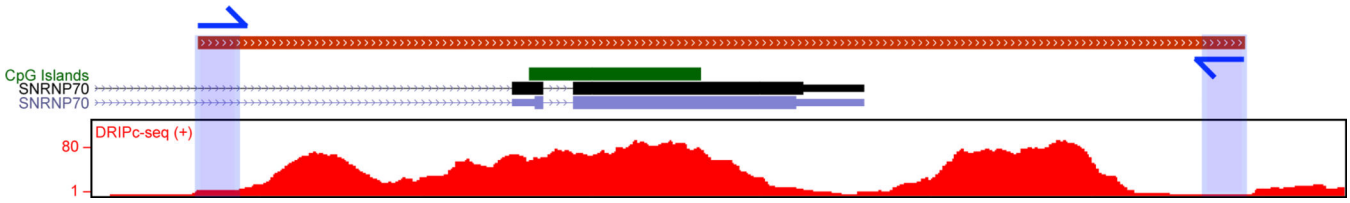

Fig. 7.

Example target region in ***SNRPN70.*** Red bar indicates possible target region to amplify. Blue arrow indicates forward and reverse primer sites flanking peaks of DRIPc-seq signal (below)