

# Isolation and sequencing of active origins of DNA replication by nascent strand capture and release (NSCR)

Dimiter Kunnev<sup>1</sup>, Amy Freeland<sup>1</sup>, Maochun Qin<sup>2</sup>, Jianmin Wang<sup>2</sup>, Steven C. Pruitt<sup>1\*</sup>

<sup>1</sup>Department of Molecular and Cellular Biology, <sup>2</sup>Department of Biostatistics and Bioinformatics, Roswell Park Cancer Institute, Buffalo, New York 14263, USA

\*Corresponding author: Steven C. Pruitt, Email: steven.pruitt@roswellpark.org

Competing interests: The authors have declared that no competing interests exist.

Abbreviations used: NSCR, nascent strand capture and release; MCM, minichromosome maintenance protein; ORC, origin recognition complex; SNS, short nascent strand; Iexo, lambda-exonuclease; CDC6, cell division cycle 6; BrdU, bromodeoxyuridine; RNase I, ribonuclease I; WGA, whole genomic DNA amplification; NGS, next generation sequencing; dNTP, deoxynucleotide triphosphate; rNTP, ribonucleotide triphosphate; MEF, mouse embryonic fibroblast; T4-PNK, T4 polynucleotide kinase

Received September 3, 2015; Revision received November 11, 2015; Accepted December 4, 2015; Published December 30, 2015

**Abstract** Nascent strand capture and release (NSCR) is a method for isolation of short nascent strands to identify origins of DNA replication. The protocol provided involves isolation of total DNA, denaturation, size fractionation on a sucrose gradient, 5'-biotinylation of the appropriate size nucleic acids, binding to a streptavidin coated magnetic beads, intensive washing, and specific release of only the RNA-containing chimeric nascent strand DNA using ribonuclease I (RNase I). The method has been applied to mammalian cells derived from proliferative tissues and cell culture but could be used for any system where DNA replication is primed by a small RNA resulting in chimeric RNA-DNA molecules.

**Keywords:** DNA replication, origin mapping, NSCR, SNS, ribonuclease I

## INTRODUCTION

DNA replication is an event of one time duplication of DNA during S-phase of the cell cycle [1]. It is multistep process that involves a large number of factors which are highly coordinated with the other cellular processes during cell cycle, differentiation and development [2-5]. The first step in DNA replication is the denaturation of DNA to form a “bubble” or “replication eye” (Fig. S1). Primase initiates synthesis of a 5'-RNA which serves as the substrate for extension by DNA polymerase. Although the size of the RNA primers may differ between organisms, the mammalian primase usually extends approximately ~11 nt (for mouse 9–11 and for human 11–14 nt) [6]. This process results in 5'-RNA-DNA-3' chimeric molecules which are the actual nascent strands [7,8]. Replication forks initiated from a given origin extend bi-directionally until merging with forks from adjacent origins. In the leading strand, nascent DNA extends continuously 5' to 3' in the direction of fork movement whereas in the lagging strand 5' to 3' synthesis of DNA occurs opposite to the movement of the replication fork. In the latter case, synthesis proceeds by repeated RNA primer synthesis by primase and short extensions of DNA, to form Okazaki fragments. Internal RNA primers are removed and Okazaki fragments are ligated together to extend the length of the lagging nascent DNA strand. Ligation of Okazaki fragments is processive such that each nascent strand becomes progressively longer while remaining centered at the origin of replication (Fig. S1). Nonetheless, they contain 5' RNA primers received from the ligation of the most recent Okazaki fragment [7,8].

## Prior methods for identifying DNA replication origins

Locations of potential DNA replication initiation sites have been inferred by identifying the locations at which components of the replication licensing complex (including ORC1, CDC6, and MCM7) are bound to chromatin using ChIP-Seq approaches [9,10]. Other methods have been developed for genome-wide identification of active DNA replication origins. These include: structure-based isolation of replication bubbles, sequence read depth analysis [11,12], strand switching of Okazaki fragments [13-16] and short nascent strand (SNS) isolation [8,17-20].

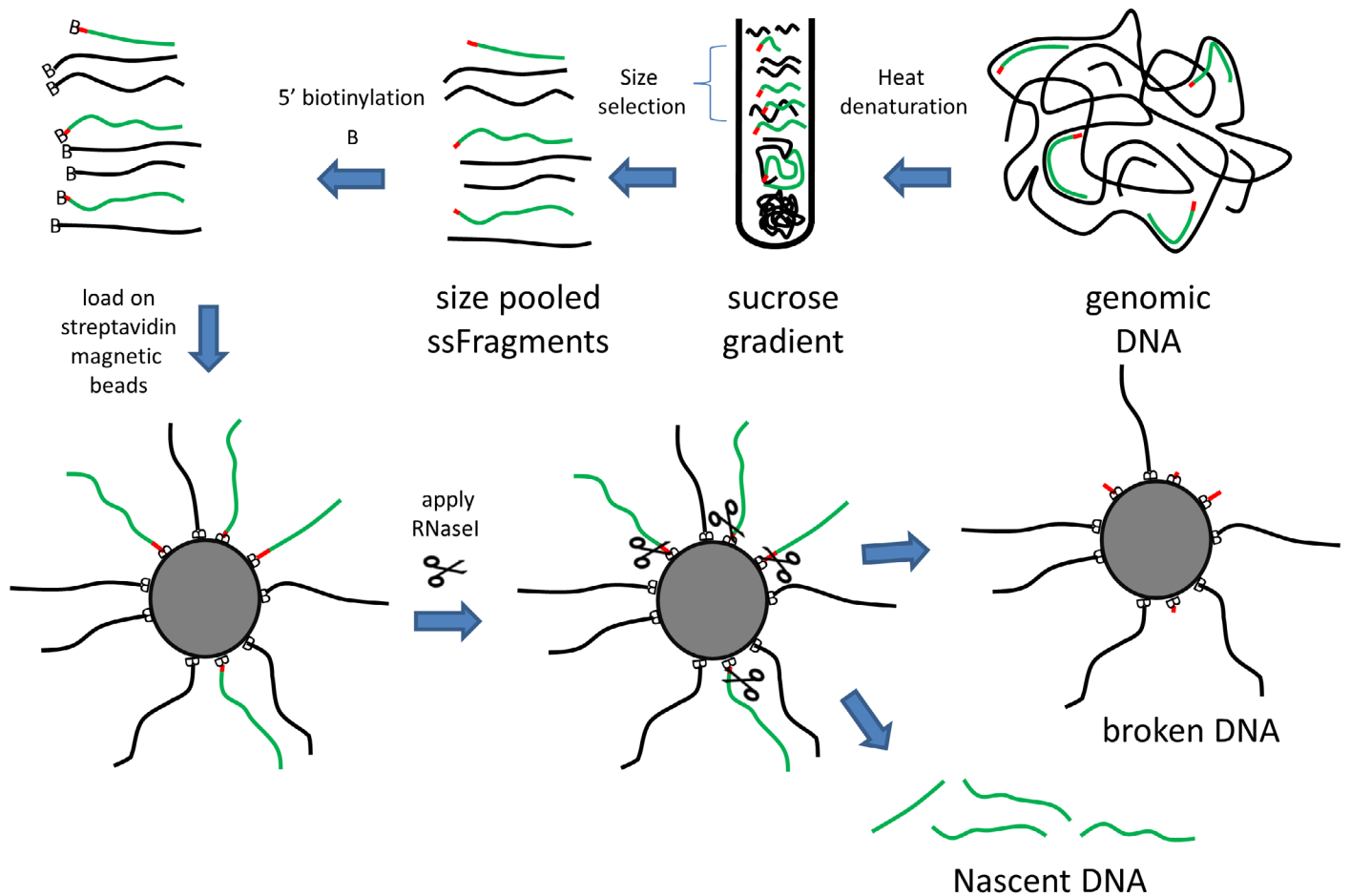
Structure-based isolation exploits the “replication bubble” present at the initiation site. Recently, a modification of the method has been applied to isolation and sequencing origins from mammalian cells where origins are first enriched based on association with nuclear matrix followed by purification of replication “bubbles” by trapping in an agarose matrix and retarding their migration during gel electrophoresis [21-23]. Limitations to this method include low resolution and sequence biases resulting from a requirement for restriction endonuclease digestion [19]. Additionally, enrichment of putative origin sequences based on attachment to the nuclear matrix may result in a distorted representation of origins across the genome. Furthermore, sequences based on structural properties other than those associated with a replication “bubble”, e.g., R-loops or repair intermediates, may also occur.

Sequence read depth approaches rely on the ability to discriminate regions of the genome that have undergone replication from those that have not by the relative copy number of different sequences represented in a sequencing library and the fact that sequences near to origins will be duplicated earlier than sequence distal to origins. In an exponentially

growing population of cells, these differences are small but recent studies suggest that bioinformatics approaches may be sufficient to identify origins based on this property [11,12].

Strand switching of Okazaki fragments results from bidirectional and semi-discontinuous replication [13-16]. To take advantage of this

property in next generation sequencing based approaches requires that knowledge of the 5'-3' orientation of isolated Okazaki fragments is retained. Such an approach has recently been applied to identification of replication origins in vaccinia virus [24].



**Figure 1. Schematic view of nascent strand capture and release (NSCR) method.** First, nascent strands (green lines with red) are released from total DNA (black lines) by heat denaturation and size selected using sucrose gradient fractionation. Gradient fractions containing fragments between 400 and 2000 nt are pooled. Fragments larger than 400 nt are large enough to avoid the Okazaki fragments and an upper size of 2000 nt insures that the sequences will lie within close proximity to their origins. The pooled fractions will contain both nascent strands and broken DNA (black lines), both of which are 5' biotinylated (indicated with B). Biotinylation is performed by first introducing a 5' thiol using T4 polynucleotide kinase (T4-PNK) with ATP $\gamma$ S as the substrate. T4-PNK transfers the thiol-group from the  $\gamma$  position of ATP $\gamma$ S to the 5'-hydroxyl terminus of single-stranded DNA and RNA. Second, the thiol is then reacted with biotin-maleimide to introduce a 5' biotin. All 5' biotin labeled fragments are loaded on streptavidin magnetic beads taking advantage of the highly specific and very strong non-covalent interaction between streptavidin and biotin. The method continues with extensive washing steps to eliminate non-specific interactions. The final step is release of the SNS by RNase I digestion of the 5' RNA primer (indicated with red color), providing a positive selection for nascent strand DNAs.

Short nascent strand (SNS) based methods for mapping replication origins are the most widely used analyses [17]. These methods depend on physical isolation and characterization of newly synthesized SNSs that reflect the beginning of replication [17,25]. Generally, DNA is isolated, denatured by heating to release nascent strands, and size fractionated to recover DNAs larger than Okazaki fragments but small enough to lie near origins of replication. The approach continues with purification of SNSs from contaminating broken DNA and characterization of the recovered sequences. The main experimental obstacle is the separation of SNSs from broken DNA of similar size, which could be a significant part of the fractionated material [17-19]. Primarily, two approaches have

been used to enrich for SNSs. In one method, replicating DNA is labeled with BrdU and selected using anti-BrdU antibodies. An advantage to this approach is a positive selection for de novo replicated DNA. A second approach for SNS enrichment is the use of lambda-exonuclease (lexo) to degrade and select against broken DNA contamination. Lexo is a 5' to 3' exonuclease that exhibits a strong preference for double stranded DNA (dsDNA) over RNA substrates. This property allows enrichment for chimeric 5'-RNA-DNA-3' fragments through preferential degradation of contaminating single strand DNAs (ssDNA) [26,27]. The activity of lexo on ssDNA is not as high as on dsDNA and therefore requires a very high concentration of the enzyme to provide significant enrichments

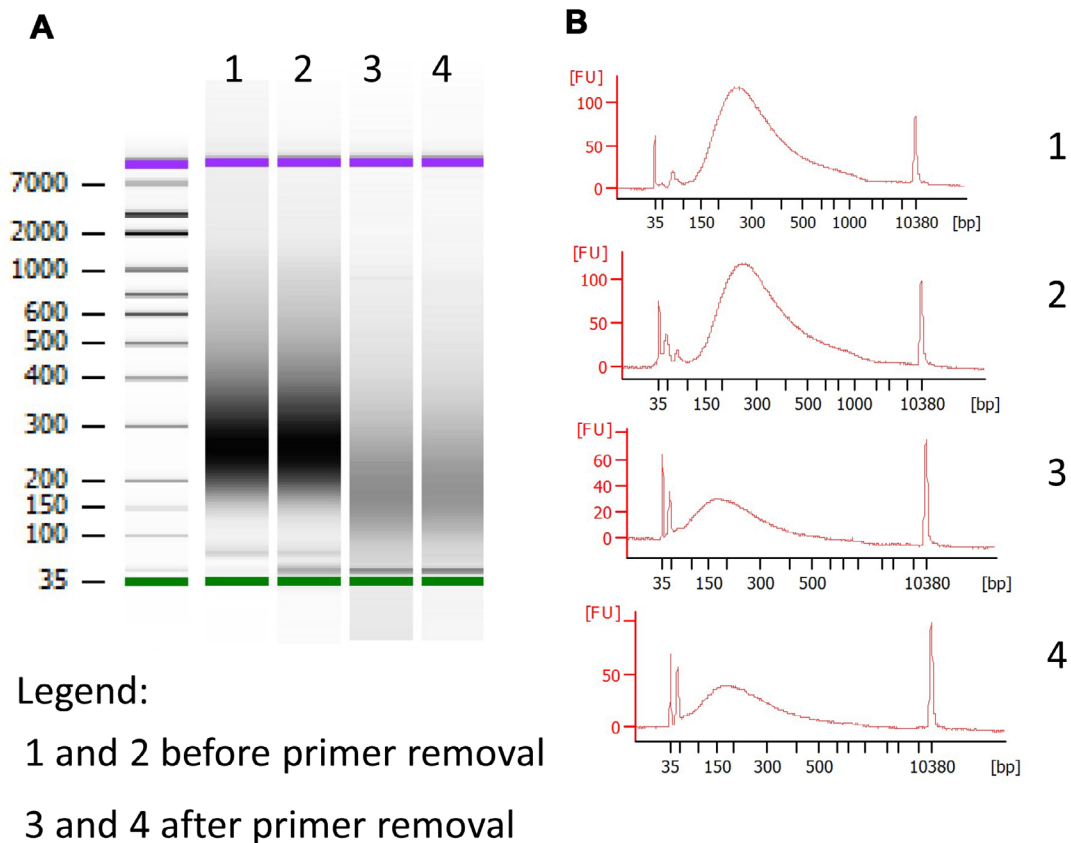
[28]. Further, lexo stalls at GC-rich sequences, can be blocked at sites of ribonucleotide incorporation, will not digest circular DNA, and can degrade RNA and hence nascent strands under some conditions, all of which could result in a sequence bias [19,26,29]. Head to head comparison of origins identified from HeLa cells by BrdU selection and lexo shows that over the same 30 Mb region BrdU selection identifies 815 origins and lexo identifies 320 origins where 49% of the origins identified by lexo are also found by BrdU selection [30].

### Development of NSCR and applications of the method and possible modifications

The nascent strand capture and release (NSCR) protocol was developed as an alternative method to enrich SNSs from size-fractionated single strand genomic DNA through a sequence-neutral, positive selection for chimeric 5'-RNA-DNA-3' molecules. Similar to the lexo method, genomic DNA is isolated, heat denatured and size fractionated to obtain the appropriately sized fragments. The material is then 5'-biotinylated using maleimide chemistry, bound to streptavidin beads, and extensively washed under conditions that minimize DNA-DNA interactions.

The method uses a final step of RNase I treatment to release the RNA containing fragments only (Fig. 1).

The developed protocol has been applied to compare SNSs from wild type and minichromosome maintenance protein 2 (MCM2) deficient mouse embryonic fibroblasts (MEFs) [31]. With minor modification NSCR could be used to examine Okazaki fragments by selecting smaller size DNA from sucrose gradients. One type of information could be achieved from Okazaki fragments is that it will mark the regions with active replication per se. Another potential application is the identification of sites of ribonucleotide incorporation into DNA. Ribonucleotide incorporation occurs at a frequency of approximately 1/2,000-1/10,000 nt in normal cells [32,33] and even a single ribonucleotide incorporated in to a deoxyribonucleotide oligomer is cleaved by RNase I (Fig. S3). Examination of sequences recovered from non-dividing cells by NSCR could allow the locations of these events to be determined. In this case the sucrose gradient should be omitted or much larger size fractions used. One caveat to this application is that ribonucleotides are replaced with deoxyribonucleotides in non-dividing cells.



**Figure 2. Bioanalyzer electrophoresis after whole genome amplification WGA.** The RNase I release fraction is used as the template for WGA, which we have performed using the SEQXE SeqPlex DNA amplification kit from Sigma-Aldrich. The workflow of the kit contains three major steps: pre-amplification in which specific primers are added to 5' and 3' ends, amplification and primer removal. **A.** Shows a bioanalyzer microgel of duplicate experiments before and after the primer removal step. **B.** Densitometry of the gels displayed in (A). As shown, the initial fragment size of 400–2000 nt is now reduced by random priming in the pre-amplification step from 150–1000 bp with a summit at ~300 bp (lines 1 and 2 before primer removal). After primer removal the size is slightly smaller 100–700 bp with a summit at ~150–200 bp (lines 3 and 4 after primer removal). This final size is suitable for Illumina sequencing.

### Experimental design (Fig. 1)

NSCR begins with purification of genomic DNA. DNA is then purified from the bulk of RNA by using TRIzol reagent because, at a later stage, RNA will compete with nascent strands during 5'-biotin labeling.

It is important to avoid using RNase to remove contaminating RNA since the 5'RNA leader in the nascent strand DNA must remain intact. DNA is re-suspended in TE buffer and heat denatured at 95–100°C for 10 min. At this point, accurate estimates of DNA concentrations can be

obtained; a total of approximately 1 mg of RNA-free DNA is recommended prior to proceeding to fractionation on sucrose gradients. For nascent strand isolation, a 5% to 30% sucrose gradient was used from which fractions containing ~400–2000 nt long DNAs are recovered and pooled. DNA of this size is sufficiently large to avoid Okazaki fragments, but small enough to lie near to origins of replication. The appropriate fractions are pooled, ethanol precipitated and washed with 70% ethanol to concentrate and desalt the DNA. After dissolving in an appropriate volume, nucleic acids are 5'-biotinylated and bound to streptavidin coated magnetic beads. Next critical step is the wash of the beads from unspecific to the biotin-streptavidin interactions. We have modified a protocol from David Wilson (published online, <http://molbio.mgh.harvard.edu/szostakweb/protocols/biotin-avidin/index.html>) for this purpose.

The final step is release of SNSs with RNase I (**Fig. 1**). RNase I is used since it exhibits little to no sequence specificity [34,35]. The RNase I released fraction constitutes the SNS product and can be assessed by PCR or amplified using whole genome amplification (WGA) in preparation for next generation sequencing. For WGA, we have used the SIGMA-ALDRICH SEQXE, SeqPlex DNA Amplification Kit following the instructions provided with primer removal. Amplified material is sheared based on the randomized 5'-primers from the kit resulting in DNA fragments approximately 200 bp in length (**Fig. 2**).

### Limitations and considerations

One limitation of the NSCR method is the efficiency. It is estimated that ~10-20 ng of SNS between 400-2000 nt is present in 600 µg of genomic DNA [8]. In comparison, NSCR routinely yields approximately 1-2 ng from 1mg total DNA; a yield of less than 10% of the theoretical calculation. Loss of SNSs may occur at several steps in the procedure including size fractionation, biotinylation, binding of biotinylated DNA to streptavidin beads, wash steps, and RNase I digestion. We observed that a significant proportion of the DNA failed to bind to the streptavidin beads suggesting that improvements in either the biotinylation or binding steps could significantly improve the yields. Due to the low yields, whole genome amplification (WGA) is necessary to achieve sufficient material for next generation sequencing (NGS). The use of WGA imposes several additional concerns. WGA results in truncation and loss of 5' to 3' orientation relative to the initial NSCR isolated SNSs. Additionally, differential amplification of different regions of the genome based on GC-content or secondary structure could alter the representation of sequences during amplification. Further, substantial bias can occur with some kits for library preparation. We obtained satisfactory results with IntegenX PrepX DNA ChIP Library Prep kit. To control for these potential sources of bias, WGA and library preparation on genomic DNA that does not contain short nascent strands can be performed. DNA from non-dividing cells or DNA from which short nascent strands have been removed (e.g., the fraction containing the largest DNA during size fraction on sucrose gradients) can be used for this purpose.

A second concern that is equally applicable to the lexo and NSCR methods is that ribonucleotides are incorporated into DNA by mechanisms independent of the activity of primase. DNA-dependent polymerases occasionally fail to distinguish between rNTPs and dNTPs (e.g. DNA polymerase delta incorporates one rNTP per every 2000 dNTPs) [32]. Inefficient removal of RNA primers from Okazaki fragments can also result in the presence of ribonucleotides at significant distances from the elongating replication fork; although RNase H2 class ribonucleases remove many of ribonucleotides from DNA *in vivo* [32, 33].

Treatments or mutations that result in an increased frequency of rNTP incorporation may result in a higher rate of inclusion of sequences that are not derived from SNSs in either lexo or NSCR isolated DNA. We tested the activity of RNase I over a single ribonucleotide incorporated into 63nt of ssDNA oligomer in same conditions as we perform NSCR. Indeed the RNase I enzymatic activity was observed on a single ribonucleotide flanked by oligo-deoxyribonucleotides (**Fig. S3**). Importantly, an incorporated ribonucleotide will be cleaved by RNase I and removed from the streptavidin beads, thereby contaminating the SNS preparation. If the sites of ribonucleotide incorporation are random in the genome, then this contamination will not be a major problem.

NSCR shows good reproducibility between experimental repeats (~90% of the largest 20% of peaks were common to two independent NSCR analyses of MEFs [31]). It may be possible to validate peaks identified by NSCR as origins by determining if peak width correlates with the size of DNAs recovered from sucrose gradients. Additionally, concordance with orthogonal methods (many of which are summarized above) would support the validity of origins identified by NSCR. Although a genome-wide head to head comparison has not been performed, a prior estimate of the number of nascent strand peaks from MEFs within a 60.4 Mbp region of mouse Chr11 has been made using lexo [28] and identified 2231 putative origins. In comparison NSCR (using a 95% cutoff value) identifies 3922 peaks in this region where 79% of the peaks identified by lexo overlap peaks identified by NSCR. However, this analysis is complicated by the differing peak widths between the two studies [31].

## MATERIALS

### Reagents

- ✓ 10% SDS
- ✓ 0.5 M EDTA
- ✓ 5 M NaCl
- ✓ 20 mg/ml proteinase K
- ✓ 1 M Tris HCl pH 8.0
- ✓ Saturated Phenol pH 8.0
- ✓ Chloroform
- ✓ Isoamyl alcohol
- ✓ Isopropanol
- ✓ Ethanol
- ✓ TRIzol Reagent (Invitrogen, Cat. # 15596-018)
- ✓ Dimethylformamide (DMF)
- ✓ Sucrose (RNase/DNase free)
- ✓ Calf intestinal phosphatase (CIP) (BioLabs, Cat. # M0290S)
- ✓ T4 polynucleotide kinase (T4 PNK) (BioLabs, Cat. # M0201S)
- ✓ ATP $\gamma$ S (Sigma-Aldrich, Cat. # A1388 1mg)
- ✓ Streptavidin, immobilized on Magnetic Beads (New England BioLabs, Cat. # S1420S)
- ✓ Biotin (Long Arm) Maleimide (Vector Laboratories, Cat. # SP-1501)
- ✓ UREA (DNase and RNase free)
- ✓ Igepal (Sigma-Aldrich, Cat. # CA-630)
- ✓ RNase I; must be protease and DNase Free, e.g., Fermentas RNase I (Cat. # EN061 10 µg/µl)

### Recipes

- ✓ Phosphate-buffered saline (1× PBS): 137 mM NaCl, 2.7 mM KCl,



- 4.3 mM Na<sub>2</sub>HPO<sub>4</sub>·7H<sub>2</sub>O, 1.4 mM KH<sub>2</sub>PO<sub>4</sub>, adjust solution to pH 7.3
- ✓ Lysis buffer: 1% SDS, 40 mM EDTA, 500 mM NaCl, 10 mM Tris HCl pH 8.0; Add proteinase K (final 0.5 mg/ml) just before using
- ✓ Chloroform-isoamyl alcohol (24:1)
- ✓ TE buffer (RNase/DNase free): 10 mM Tris-HCL pH 8.0, 1 mM EDTA
- ✓ 5% and 30% sucrose buffers: sucrose 5% or 30%, 1 mM EDTA, 10 mM TrisHCl pH 8.0; 300 mM NaCl
- ✓ 5'-EndTag Nucleic Acid Labeling System (Vector Laboratories, Cat. # MB-9001)

**OPTIONAL:** If you do not use the Vector Laboratories kit you may prepare the components necessary for 5'-biotin labeling as follows:

- Universal buffer: a 10× stock could be prepared by mixing equal volumes of 10× Calf Intestinal Phosphatase (CIP) buffer and 10× polynucleotide kinase (T4) buffer.
- ATPγS (Sigma-Aldrich, Cat. # A1388): 1 mg dissolved in 769 μl ddH<sub>2</sub>O (final concentration: 1.3 μg/μl). It can be stored at -80°C for a long period of time.
- ✓ Biotin maleimide 12 mg dissolved in 312 μl Dimethylformamide (DMF) according to the Vector Laboratories kit instructions
- ✓ Buffer A (high salt): 1 M NaCl; 20 mM Tris pH 7.5; 5 mM EDTA; 0.1% Igepal
- ✓ Buffer B (low salt): TE with 0.1% Igepal
- ✓ Buffer C (UREA containing): 4 M Urea; 10 mM Tris pH 8.0; 1 mM EDTA; 0.1% Igepal
- ✓ 20× Saline-sodium citrate buffer (20× SSC): 3 M NaCl, 300 mM

Sodium citrate adjust the pH to 7.0

- ✓ RNase I Buffer (RNB): 10 mM Tris pH 8.0; 100 mM NaCl; 5 mM EDTA

**NOTE:** RNase and DNase free solutions and working conditions are important throughout the protocol. Use commercially available RNase and DNase free water and buffers. Alternatively, autoclaved buffers and entire glassware may be used. Use new tubes with proven RNase DNAase free status. We did not test DEPC-treated water for NSCR.

## Equipment

- ✓ Ultracentrifuge
- ✓ SW28.1 Beckman ultracentrifuge rotor
- ✓ 17-ml tubes for SW28.1 Beckman rotor
- ✓ Glass tubes rated for use above 3000 × g for the TRIzol procedure

**OPTIONAL:** gradient maker for preparing sucrose gradients, alternatively sucrose gradients can be prepared manually as described in **Figure S2**.

- ✓ 6-Tube Magnetic Separation Rack (New England BioLabs, Cat. # S1506S)
- ✓ Falcon tubes 50 ml and 15 ml
- ✓ Eppendorf tubes 1.5 ml
- ✓ Heating blocks for 65°C and 70°C

## PROCEDURE

### Purification of genomic DNA from cultured cells by SDS lysis buffer and protease K (Timing: 1 Day) (Described in Supplementary Information)

**NOTE:** Cells or tissue: Cultured cells or tissues can be used as starting material but should contain proliferating cells and be sufficient to provide at least 1 mg of purified DNA.

### Split DNA from RNA by using TRIzol Reagent (Timing 1 Day) (Described in Supplementary Information)

**NOTE:** It important to follow the TRIzol protocol until you collect the middle white rubber-like DNA fraction.

**CAUTION:** Do not follow the rest of the TRIzol protocol as it uses treatment with NaOH in order to dissolve the DNA. Alkali conditions will destroy the RNA primers required for SNS capture.

### Size fractionation of nascent DNA + DNA fragments by 16 ml 5%-30% sucrose gradient (Timing 1 Day) (Described in Supplementary Information)

**NOTE:** A typical yield is ~100–200 μg of sized (~400–2000 nt) pooled DNA from starting 1 mg RNA free DNA. Proceed directly with biotinylation or keep it at -20°C for up to one week.

### 5'-labeling nascent strand DNA + DNA fragments with biotin for one reaction (for up to 0.6 nmol 5' ends - can be scaled for larger amounts) (Timing 4 h) (Described in Supplementary Information)

**NOTE:** 5'-labeling is well described in the Vector Laboratories kit: 5'-EndTag Nucleic Acid Labeling System (Cat. # MB-9001). We modified the protocol slightly with extended time for different incubations from 30 min to 1 h. You may need to scale up the volumes based on your initial DNA amount.

## Capture nascent DNA + DNA fragments on streptavidin magnetic beads (Timing ~4 h)

1. Pipet the calculated volume of streptavidin coated magnetic beads into a 1.5 ml tube using a cut tip. Wash the beads by adding 1 ml Buffer A, rinse the tip in the buffer to collect any adhering beads, apply the magnet for 3–5 min until the supernatant is clear and the beads are all collected into a spot adjacent to the magnet. Remove the supernatant taking care not to disturb the beads.

**NOTE:** The beads concentration is 4  $\mu\text{g}/\mu\text{l}$  and the binding capacity is 1 pmol of nucleic acid 5'ends/1  $\mu\text{g}$  of beads or 4 pmols/ $\mu\text{l}$  or 4 nmols/ml. Assuming that 5'ends are 100% labeled, 100 ng of 500 nt ssDNA (or RNA) has 0.6 pmol of 5'ends (see Supplemental Information section E). Therefore theoretically 1  $\mu\text{l}$  of beads is enough to bind ~600 ng of ssDNA of ~500 nt in length. Starting with ~1–2 mg of DNA (RNA free) for sucrose gradient centrifugation, the typical yield we observe after biotinylation of 400–2000 base fractions is around 100–200  $\mu\text{g}$  of biotinylated DNA, equivalent to 0.6–1.2 nmols of 5' ends. (The yield will depend on a number of factors including the proportion of dividing cells in the starting sample and the level of DNA breakage during isolation.) ~150  $\mu\text{l}$  of beads is required per 100  $\mu\text{g}$  5' biotinylated DNA. Using these parameters, the bead volume required to provide an excess of binding capacity should be calculated. If more than one sample is being prepared, use an equal volume of beads for all reactions ensuring that the amount is sufficient for the sample with the greatest amount of DNA. It is important not to overload the bead capacity.

2. Add 300  $\mu\text{l}$  Buffer A to the beads and mix well.
3. Add 300  $\mu\text{l}$  Buffer A to the biotinylated fraction (it is step 19 from Supplementary Information protocol) and mix well.
4. Combine the beads (300  $\mu\text{l}$ , step 2) with biotinylated sample (350  $\mu\text{l}$ , step 3) and mix well by finger flicking.
5. Incubate for 60 min at room temperature with shaking or rocking.
6. Apply the magnet as described above.
7. Collect the supernatant material as non-bound ~650  $\mu\text{l}$  flow-through (**Fraction F**).
8. Wash twice with 1 ml Buffer A at room temperature.
9. Wash twice with 1 ml Buffer B at room temperature.
10. Wash three times with 1 ml Buffer C in a heating block at +70°C for 5 min.
11. Wash three times with 1 ml 20 $\times$  SSC in a heating block at +70°C for 10 min.
12. Switch off the heating block and perform the rest of the wash steps as the temperature declines.
13. Wash with 1 ml TE.
14. Add 1 ml RNB re-suspend well and leave it on the heating block for 30 min. Apply the magnet and remove the buffer.

**NOTE:** Washing the beads to remove non-specifically bound DNA is the most important step of NSCR. The intensive treatment with high salt, low salt, heating at +70°C in 4 M Urea and 20 $\times$  SSC reduces the background due to non-specific DNA-DNA and DNA-bead interactions.

**CAUTION:** All wash steps need to be performed with the appropriate buffer and at the described temperature using 1 ml of buffer. Finger-tap the tubes to be sure that the beads are well dispersed then incubate at the temperatures described. Apply the magnet for sufficient time to clear the supernatant of beads prior to removing it. The volume of the beads shrinks a bit during the washes. If un-biotinylated DNA is used, the beads will be lost much more rapidly, possibly completely, by the end of the wash cycles.

15. Add 400  $\mu\text{l}$  RNB Buffer to the beads, re-suspend well and incubate for 15 min at room temperature, apply the magnet and collect the supernatant as the mock RNase I wash control (**Fraction W**). The wash control serves to estimate the non-specific background in the absence of RNase I release of SNS.

## Release the Nascent DNA with RNase I (Timing 1 h)

16. Make a mixture (or master mix) of 1:500 (= 0.02 U/ $\mu\text{l}$ ) dilution of RNase I into RNB buffer.
17. Add 200  $\mu\text{l}$  of the RNase I mixture to the beads. Mix well, by finger tapping or pipetting up and down several times. Incubate 15 min at room temperature, not longer.
18. Apply the magnet and collect the supernatant.
19. Wash the beads by adding 200  $\mu\text{l}$  of RNB to the beads, mix well and apply the magnet again to collect the supernatant.

20. Combine first 200  $\mu\text{l}$  with this second 200  $\mu\text{l}$  which together is the RNase I release fraction, **Fraction R**.

**CAUTION:** Fractions **F**, **W** and **R** may contain residual beads, which could increase the background. In order to minimize any contamination from left over beads, proceed to next step.

21. Spin down the fractions **F**, **W** and **R** in a microfuge at maximum rpm for 5 min. Collect the entire volume into a new tube except the last few microliter which may contain the residual beads if any.

**PAUSE POINT:** The fractions can be stored at  $-20^{\circ}\text{C}$ .

**NOTE:** In most cases beads will not be visible, but may nonetheless be present. Store all fractions and the beads at  $-20^{\circ}\text{C}$ . In addition if you need to assess the DNA left remaining bound to the beads: Treat the beads with 0.4 M NaOH (100  $\mu\text{l}$ ) for 5 min at room temperature, neutralize with 100  $\mu\text{l}$  of 0.4 M HCl, and spin for 1 min at maximum speed. Collect the supernatant as the bound or column fraction (Fraction C).

### Concentrating nascent strands and preparing for whole genome amplification

22. Perform a phenol extraction of fractions **W** and **R** using 100  $\mu\text{l}$  phenol. Re-extract the phenol with an additional 200  $\mu\text{l}$  TE and combine the aqueous phases. Precipitate by addition of 5  $\mu\text{l}$  glycogen, 1/10 volume of 3 M NaAc (mix well) and one volume of isopropanol.

**PAUSE POINT:** DNA can be stored for long periods of time in isopropanol at  $-20^{\circ}\text{C}$ .

23. Spin at 14,000 rpm (maximum speed in a microfuge) for at least 45 min at room temperature, look for the pellet, and then discard supernatant. Wash with 70% EtOH 3 times. Perform the wash steps with caution taking care not to lose the pellet.

**PAUSE POINT:** DNA can be stored for long periods of time in ethanol at  $-20^{\circ}\text{C}$ .

24. Dry and re-suspend in 20–30  $\mu\text{l}$  TE.

**NOTE:** Fractions W and R are ready for PCR or for whole genome amplification (WGA). The amount of DNA in these fractions is in the range of picograms per microliter and is difficult to measure directly by absorbance or pico-green. Pico-green interacts with dsDNA but does not detect single stranded nascent DNAs efficiently except for the regions with secondary structure. For higher precision of pico-green measurement ssDNA can be converted to dsDNA using the Klenow fragment of DNA polymerase I. Other methods for measuring the amount of DNA were not tested.

25. Perform a quality control with 50  $\mu\text{l}$  PCR reaction(s) at  $60^{\circ}\text{C}$  annealing temperature with 27–29 cycles for a known strong peak using only 2  $\mu\text{l}$  from fraction W (Wash control) and from R (RNase I release). It is important to observe a significant enrichment of the band signal for fraction R over W which will be an indication of RNase I dependent release into fraction R. For the mouse genome we used primers: forward F-Chr11\_108873.3 CCTGCCCTTCACAAAGAAAA and reverse R-Chr11\_108873.3 GTGGAGAAGGGTTCCATGTG. High quality NSCR preparations should demonstrate at least a 3–10 times brighter band for the R fraction versus the W fraction. It is typical however for the W fraction to give some background amplification. Different sites may demonstrate different levels of enrichment.

### Whole genome amplification (WGA)

26. We have used the WGA kit SeqPlex Enhanced DNA Amplification Kit Sigma “SEQXE-10RXN” plus primer removal step. Other methods or kits for whole genome amplification were not tested. The protocol provided with the kit was followed without deviation except that 29 cycles were used for the amplification (**Fig 2**).

### Validation of NSCR preparation

27. Following the primer removal step in 26, free primers are removed from the sample using a PCR clean-up kit (e.g., GenElute or PureLink PCR purification kit cat No K3100-01) with elution in 50  $\mu\text{l}$ . ~1–2  $\mu\text{l}$  of the eluate is run on a bioanalyser (High sensitivity DNA assay) to estimate the size and the amount of amplified fragments (**Fig 2**). A preparation is considered good if you observe a peak summit at ~200 bp and a total amount of at least 50 ng.

## Sequencing library preparation

- Library preparation is performed using an IntegenX PrepX DNA ChIP library prep kit. We also tested the NEB-Next ChIP Seq Library Prep kit, however a relatively high GC bias was observed. Other library preparation kits were not tested.

## Sequencing recommendations

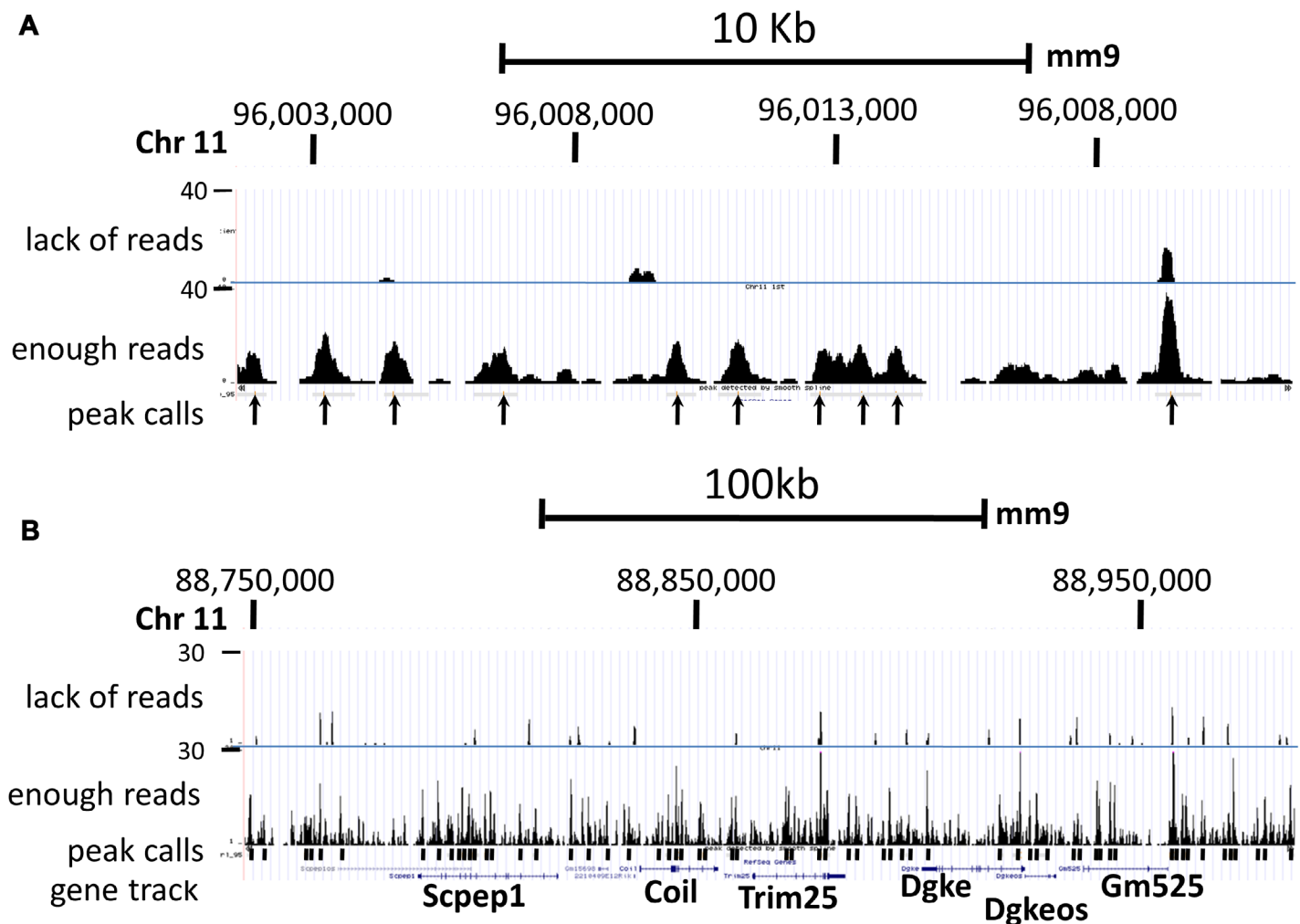
We recommend pair-end 100 bp for the sequencing for increased mappability. A total of ~250–300 million uniquely mapped sequences yields good coverage of mammalian genomes (Fig. 3).

## Sample data location

Data can be found at GEO with access number of GSE64933 and NCBI BioProject database under accession number PRJNA258088.

## Bioinformatic analysis of nascent strand sequences

A variety of approaches to the bioinformatic analysis of nascent strand peaks may be used. The approach we have taken is provided in the Supplemental Information Section including : specific software and hardware requirements, demonstrations of command line operations, expected results of the operations, peak calling/putative replication origin identification, selection of windows with potential peaks, peak detection in selected windows, exact replication origin identification, coverage normalization and difference wiggle files generation between experimental and control samples, and sequence feature searches for replication origins.



**Figure 3. Examples of wiggle files plotted on the UCSC genome browser showing sufficient and insufficient reads in two different scales.** NSCR purified SNS were WGA amplified and sequenced by Illumina sequencing. The resulting fastq files were aligned to the mm9 genome build and wiggle files were generated. It is shown few tracks: with insufficient reads (~0.6 fold coverage), with sufficient reads (~4.5 fold coverage), RefSeq genes and peaks detected with described algorithm above at a random location on chromosome 11. Panel (A) visualized all tracks on 10 kb scale and panel (B) on 100 kb scale.



**Table 1. Troubleshooting.**

Step	Problem	Causes	Suggestions
Suppl II D. Step 5	<ul style="list-style-type: none"> <li>Fractionation gel shows second bands with lower molecular weight.</li> <li>On the gel the high molecular weight DNA is present in the lower fractions contaminating my fractions of choice with longer DNA.</li> </ul>	<ul style="list-style-type: none"> <li>The second bands are RNA, which was not removed during the TRIzol step.</li> <li>Most likely this results from poor re-suspension of DNA after heating and before loading on the sucrose gradient (e.g. clumps of DNA were present).</li> <li>Alternatively, the sucrose gradient may have been overloaded.</li> </ul>	<ul style="list-style-type: none"> <li>Performing a TRIzol step to split DNA from RNA was not successful. Collect only the interphase white pellet without aqueous phase. If you have a very well formed tight white pellet, most likely you will obtain a good separation of DNA and RNA.</li> <li>The DNA need to be well dissolved after heating.</li> <li>Do not load more than 350-400 mg DNA per 18 ml gradient.</li> </ul>
8-13	<ul style="list-style-type: none"> <li>Too few beads remain after washing.</li> </ul>	<ul style="list-style-type: none"> <li>The biotinylation level is too low.</li> <li>Your input DNA is not as much as you expected.</li> </ul>	<ul style="list-style-type: none"> <li>Make sure that you perform the 5'biotinylation correctly with fresh (proven) reagents.</li> <li>Use more input DNA.</li> </ul>
25	<ul style="list-style-type: none"> <li>PCR was performed on fractions W and R for a proven origin in order to demonstrate origin enrichment. However, no significant enrichment in the R fraction over the W fraction was seen.</li> </ul>	<ul style="list-style-type: none"> <li>Wash step was not sufficient.</li> <li>The representation of the proven origin as part of SNS after NSCR is so low that is difficult to detect using a reasonable (up to 32) number of PCR cycles.</li> </ul>	<ul style="list-style-type: none"> <li>Be sure that you perform the washing conditions as it is described. Make fresh washing buffers especially the urea containing buffer (buffer C).</li> <li>Perform PCR on another site which usually demonstrates stronger signal.</li> </ul>

## ANTICIPATED RESULTS

The result described in **Figure 2** demonstrates a sufficiently well represented distribution of amplified SNS on a microgel bioanalyzer. **Figure 2** shows that the overall the size of the summit in the gel is ~300 bp before primer removal step and ~200 bp after primer removal. If no significant material is detected after the primer removal step origin sequences will not be well represented in NGS data from the sample. **Figure 3** shows wiggle track files aligned to the mm9 genome and plotted with the UCSC web browser for the same region and at the same resolution. Approximately 20 kb and 200 kb regions are shown and include scale, NSCR tracks with insufficient or sufficient reads, peak track and gene track as indicated in the figure.

## Acknowledgments

This work was supported by grants from the NIH (RO1 CA130995, RO1 AG041854) to SCP.

## References

- Machida YJ, Hamlin JL, Dutta A (2005) Right place, right time, and only once: replication initiation in metazoans. *Cell* 123: 13-24. doi: [10.1016/j.cell.2005.09.019](https://doi.org/10.1016/j.cell.2005.09.019). PMID: 16213209
- Bell SP, Dutta A (2001) DNA replication in eukaryotic cells. *Annu Rev Biochem* 71: 333-374. doi: [10.1146/annurev.biochem.71.110601.135425](https://doi.org/10.1146/annurev.biochem.71.110601.135425). PMID: 12045100
- DePamphilis ML, Blow JJ, Ghosh S, Saha T, Noguchi K, et al. (2006) Regulating the licensing of DNA replication origins in metazoa. *Curr Opin Cell Biol* 18: 231-239. doi: [10.1016/j.ccb.2006.04.001](https://doi.org/10.1016/j.ccb.2006.04.001). PMID: 16650748
- DePamphilis ML (2005) Cell cycle dependent regulation of the origin recognition complex. *Cell Cycle* 4: 70-79. PMID: 15611627
- Fragkos M, Ganier O, Coulombe P, Méchali M (2015) DNA replication origin activation in space and time. *Nat Rev Mol Cell Biol* 16: 360-374. doi: [10.1038/nrm4002](https://doi.org/10.1038/nrm4002). PMID: 25999062
- Frick DN, Richardson CC (2001) DNA primases. *Annu Rev Biochem* 70: 39-80. doi: [10.1146/annurev.biochem.70.1.39](https://doi.org/10.1146/annurev.biochem.70.1.39). PMID: 11395402
- Berg JM, Tymoczko JL, Stryer L (2002) *Biochemistry*. 5th edition. New York:

W H Freeman. Chapter 27, DNA Replication, Recombination, and Repair; pp. 745-780.

- Hamlin JL, Mesner LD, Dijkwel PA (2010) A winding road to origin discovery. *Chromosome Res* 18: 45-61. doi: [10.1007/s10577-009-9089-z](https://doi.org/10.1007/s10577-009-9089-z). PMID: 19859818
- Dellino GI, Cittaro D, Piccioni R, Luzzi L, Banfi S, et al. (2012) Genome-wide mapping of human DNA-replication origins: levels of transcription at ORC1 sites regulate origin selection and replication timing. *Genome Res* 23: 1-11. doi: [10.1101/gr.142331.112](https://doi.org/10.1101/gr.142331.112). PMID: 23187890
- Sugimoto N, Maehara K, Yoshida K, Yasukouchi S, Osano S, et al. (2015) Cdt1-binding protein GRWD1 is a novel histone-binding protein that facilitates MCM loading through its influence on chromatin architecture. *Nucleic Acids Res* 43: 5898-5911. doi: [10.1093/nar/gkv509](https://doi.org/10.1093/nar/gkv509). PMID: 25990725
- Xu J, Yanagisawa Y, Tsankov AM, Hart C, Aoki K, et al. (2012) Genome-wide identification and characterization of replication origins by deep sequencing. *Genome Biol* 13: doi: [10.1186/gb-2012-13-4-r27](https://doi.org/10.1186/gb-2012-13-4-r27). PMID: 22531001
- Müller CA, Hawkins M, Retkute R, Malla S, Wilson R, et al. (2013) The dynamics of genome replication using deep sequencing. *Nucleic Acids Res* 42: doi: [10.1093/nar/gkt878](https://doi.org/10.1093/nar/gkt878). PMID: 24089142
- Hay RT, DePamphilis ML (1982) Initiation of SV40 DNA replication *in vivo*: location and structure of 5' ends of DNA synthesized in the ori region. *Cell* 28: 767-779. PMID: 6178514
- Kohara Y, Tohdoh N, Jiang XW, Okazaki T (1985) The distribution and properties of RNA primed initiation sites of DNA synthesis at the replication origin of *Escherichia coli* chromosome. *Nucleic Acids Res* 13: 6847-6866. PMID: 2414732
- Gerbi SA, Bielinsky AK (1997) Replication initiation point mapping. *Methods* 13: 271-280. doi: [10.1006/meth.1997.0526](https://doi.org/10.1006/meth.1997.0526). PMID: 9441853
- Burhans WC, Vassilev LT, Caddle MS, Heintz NH, DePamphilis ML (1990) Identification of an origin of bidirectional DNA replication in mammalian chromosomes. *Cell* 62: 955-965. PMID: 2393905
- Gilbert DM (2010) Evaluating genome-scale approaches to eukaryotic DNA replication. *Nat Rev Genet* 11: 673-684. doi: [10.1038/nrg2830](https://doi.org/10.1038/nrg2830). PMID: 20811343
- Gerbi SA (2005) Mapping origins of DNA replication in eukaryotes. *Methods Mol Biol* 296: 167-180. PMID: 15576931
- Urban JM, Foulk MS, Casella C, Gerbi SA (2015) The hunt for origins of DNA replication in multicellular eukaryotes. *F1000Prime Reports*. eCollection Review 7: 1-16. doi: [10.12703/P7-30](https://doi.org/10.12703/P7-30). PMID: 25926981
- Hyrien O (2015) Peaks cloaked in the mist: the landscape of mammalian replication origins. *J Cell Biol* 208: 147-160. doi: [10.1083/jcb.201407004](https://doi.org/10.1083/jcb.201407004). PMID: 25601401
- Mesner LD, Valsakumar V, Cieslik M, Pickin R, Hamlin JL, et al. (2013) Bubble-seq analysis of the human genome reveals distinct chromatin-mediated mechanisms for regulating early- and late-firing origins. *Genome Res* 23: 1774-1788. doi: [10.1101/gr.155218.113](https://doi.org/10.1101/gr.155218.113). PMID: 23861383

22. Mesner LD, Valsakumar V, Karnani N, Dutta A, Hamlin JL, et al. (2010) Bubble-chip analysis of human origin distributions demonstrates a genomic scale significant clustering into zones and significant association with transcription. *Genome Res* 21: 377-389. doi: [10.1101/gr.111328.110](https://doi.org/10.1101/gr.111328.110). PMID: 21173031
23. Mesner LD, Hamlin JL (2009) Isolation of restriction fragments containing origins of replication from complex genomes. *Methods Mol Biol* 521: 315-328. doi: [10.1007/978-1-60327-815-7\\_17](https://doi.org/10.1007/978-1-60327-815-7_17). PMID: 19563114
24. Senkevich TG, Bruno D, Martens C, Porcella SF, Wolf YI, et al. (2015) Mapping vaccinia virus DNA replication origins at nucleotide level by deep sequencing. *Proc Natl Acad Sci U S A* 112: 10908-10913. doi: [10.1073/pnas.1514809112](https://doi.org/10.1073/pnas.1514809112). PMID: 26286988
25. Vassilev L, Johnson EM (1990) An initiation zone of chromosomal DNA replication located upstream of the c-myc gene in proliferating HeLa cells. *Mol Cell Biol* 10: 4899-4904. PMID: 2201906
26. Foulk MS, Urban JM, Casella C, Gerbi SA (2015) Characterizing and controlling intrinsic biases of Lambda exonuclease in nascent strand sequencing reveals phasing between nucleosomes and G-quadruplex motifs around a subset of human replication origins. *Genome Res* 25: 1-11. doi: [10.1101/gr.183848.114](https://doi.org/10.1101/gr.183848.114). PMID: 25695952
27. Bielinsky AK, Gerbi SA (1998) Discrete start sites for DNA synthesis in the yeast ARS1 origin. *Science* 279: 95-98. PMID: 9417033
28. Cayrou C, Coulombe P, Vigneron A, Stanojcic S, Ganier O, et al. (2011) Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome Res* 21: 1438-1449. doi: [10.1101/gr.121830.111](https://doi.org/10.1101/gr.121830.111). PMID: 21750104
29. Perkins TT, Dalal RV, Mitsis PG, Block SM (2003) Sequence-dependent pausing of single lambda exonuclease molecules. *Science* 301: 1914-1918. doi: [10.1126/science.1088047](https://doi.org/10.1126/science.1088047). PMID: 12947034
30. Karnani N, Taylor CM, Malhotra A, Dutta A (2009) Genomic study of replication initiation in human chromosomes reveals the influence of transcription regulation and chromatin structure on origin selection. *Mol Biol Cell* 21: 393-404. doi: [10.1091/mbc.E09-08-0707](https://doi.org/10.1091/mbc.E09-08-0707). PMID: 19955211
31. Kunnev D, Freeland A, Qin M, Leach RW, Wang J, et al. (2015) Effect of minichromosome maintenance protein 2 deficiency on the locations of DNA replication origins. *Genome Res* 25: 558-569. doi: [10.1101/gr.176099.114](https://doi.org/10.1101/gr.176099.114). PMID: 25762552
32. Clausen AR, Zhang S, Burgers PM, Lee MY, Kunkel TA (2012) Ribonucleotide incorporation, proofreading and bypass by human DNA polymerase  $\delta$ . *DNA Repair (Amst)* 12: 121-127. doi: [10.1016/j.dnarep.2012.11.006](https://doi.org/10.1016/j.dnarep.2012.11.006). PMID: 23245697
33. Potenski CJ, Klein HL (2014) How the misincorporation of ribonucleotides into genomic DNA can be both harmful and helpful to cells. *Nucleic Acids Res* 42: 10226-10234. doi: [10.1093/nar/gku773](https://doi.org/10.1093/nar/gku773). PMID: 25159610
34. Spahr PF, Hollingworth BR (1961) Purification and Mechanism of Action of Ribonuclease from *Escherichia coli* Ribosomes. *J Biol Chem* 236: 823-831.
35. Meador J 3rd, Cannon B, Cannistraro VJ, Kennell D (1990) Purification and characterization of *Escherichia coli* RNase I. Comparisons with RNase M. *Eur J Biochem* 187: 549-553. PMID: 2406134

## Supplementary information

**Figure S1.** Supplementary introduction figure.

Supplementary methods

Bioinformatics and analysis of nascent strand sequences

Supplementary references

Supplementary information of this article can be found online at <http://www.jbmethods.org/jbm/rt/suppFiles/92>.