NEBNext Direct: A Novel, Rapid, Hybridization-Based Approach for the Capture and Library Conversion of Genomic Regions of Interest

Amy B. Emerman,[1] Sarah K. Bowman,[1] Andrew Barry,[2] Noa Henig,[1] Kruti M. Patel,[1] Andrew F. Gardner,[2] and Cynthia L. Hendrickson[1,3]

[1]Directed Genomics, Inc., Ipswich, Massachusetts
[2]New England Biolabs, Ipswich, Massachusetts
[3]Corresponding author (*hendrickson@directedgenomics.com*)

Next-generation sequencing (NGS) is a powerful tool for genomic studies, translational research, and clinical diagnostics that enables the detection of single nucleotide polymorphisms, insertions and deletions, copy number variations, and other genetic variations. Target enrichment technologies improve the efficiency of NGS by only sequencing regions of interest, which reduces sequencing costs while increasing coverage of the selected targets. Here we present NEBNext Direct®, a hybridization-based, target-enrichment approach that addresses many of the shortcomings of traditional target-enrichment methods. This approach features a simple, 7-hr workflow that uses enzymatic removal of off-target sequences to achieve a high specificity for regions of interest. Additionally, unique molecular identifiers are incorporated for the identification and filtering of PCR duplicates. The same protocol can be used across a wide range of input amounts, input types, and panel sizes, enabling NEBNext Direct to be broadly applicable across a wide variety of research and diagnostic needs. © 2017 by John Wiley & Sons, Inc.

Keywords: in-solution hybridization • Illumina library preparation • next-generation sequencing • NGS library preparation • NGS target capture • PCR enrichment • target enrichment

How to cite this article:
Emerman, A. B., Bowman, S. K., Barry, A., Henig, N., Patel, K. M., Gardner, A. F., & Hendrickson, C. L. (2017). NEBNext direct: A novel, rapid, hybridization-based approach for the capture and library conversion of genomic regions of interest. *Current Protocols in Molecular Biology*, *119*, 7.30.1–7.30.24. doi: 10.1002/cpmb.39

INTRODUCTION

Targeted sequencing technologies facilitate massively parallel sequencing of user-defined genomic regions (Kozarewa, Armisen, Gardner, Slatko, & Hendrickson, 2015). These approaches focus data collection solely on the targets of interest, as opposed to the whole genome. The end result is extensive sequencing coverage of desired regions, a lower cost of sequencing, and in the case of human patients, elimination of ethical concerns arising from the collection of off-target genetic information.

Here we describe the NEBNext Direct® target enrichment protocol for Illumina® next-generation sequencing (Fig. 7.30.1). This hybridization-based approach facilitates production of sequencing-ready libraries from fragmented DNA in as few as 7 hr. Fragmented DNA is first hybridized to biotinylated oligonucleotide baits in a short, 90-min

Current Protocols in Molecular Biology 7.30.1–7.30.24, July 2017
Published online July 2017 in Wiley Online Library (wileyonlinelibrary.com).
doi: 10.1002/cpmb.39

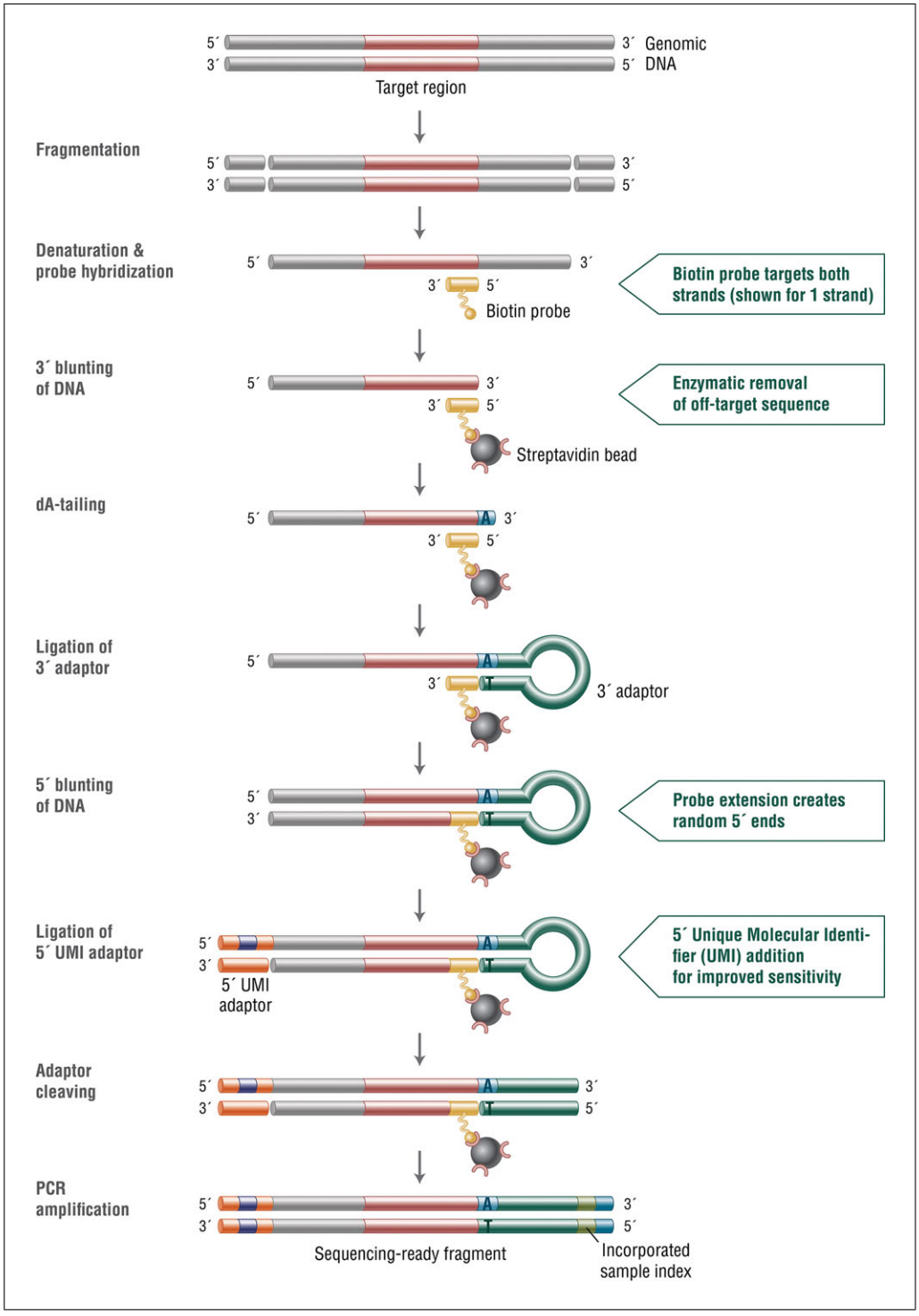

Figure 7.30.1 Schematic of the NEBNext Direct protocol. Reproduced from *https://www.neb.com/nebnext-direct/nebnext-direct-for-target-enrichment*.

hybridization step. The baits define the 3′ end of each target and are designed to capture both strands of the input DNA. Immediately after hybridization, the bait-target hybrids are captured on magnetic streptavidin beads, which allow for simple, magnet-based retention of the bead-bound DNA for solution exchanges and post-reaction washes in the subsequent enzymatic steps. After capture, enzymatic removal of off-target sequences focuses library construction only on regions of interest; thus, off-target regions make an extremely limited contribution to the final library. Next, Illumina-compatible 3′ and 5′ adaptors are added to the DNA targets in separate steps that follow dA-tailing or blunting

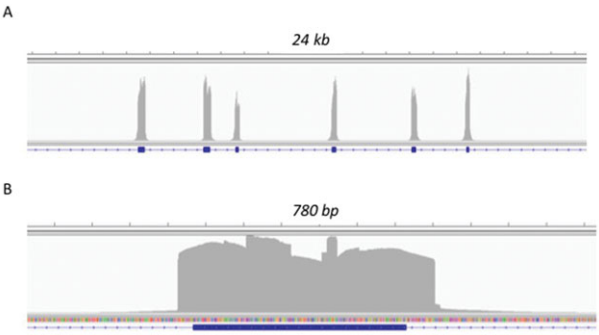

Figure 7.30.2 Representative target coverage plots for NEBNext Direct libraries visualized in the Integrative Genomics Viewer (IGV) (Robinson et al., 2011; Thorvaldsdóttir, Robinson, & Mesirov, 2013). (**A**) A 24-kb view of several exons. (**B**) A 780-bp view of a single exon.

of each end, respectively. Finally, unwanted synthetic DNA sequences are cleaved in preparation for amplification. Following amplification, the libraries are cleaned up and size selected using SPRI beads. After proper quantification and quality-control analysis, the resulting Illumina-compatible libraries are used for high-throughput sequencing.

Multiple libraries prepared from different sources of DNA or RNA may be assayed in parallel, as each library sample receives an 8-base sample index. NEBNext Direct currently provides up to 120 individual sample index barcodes to enable multiplexing of many samples in one sequencing run (dual-index barcodes, which include sample barcodes in both the i5 and i7 positions to provide a greater number of total indexes, are available on a custom basis). Additionally, each individual captured molecule within a sample receives a 12-base unique molecular identifier (UMI). The UMIs facilitate bioinformatic identification of both PCR duplicates and sequencing errors, thus increasing the sensitivity and fidelity of variant detection.

The NEBNext Direct method features strengths from both PCR-based and hybridization-based target-enrichment approaches, and contains several features that distinguish it from other target-enrichment technologies. First, the NEBNext Direct approach captures targets of interest prior to their incorporation into a library, thereby focusing the resources for library construction only on regions of interest. Second, the sequencing read coverage of targeted genomic regions from NEBNext Direct libraries have highly defined beginning and end points, resembling the target coverage profiles of PCR-based enrichment methods (Fig. 7.30.2). This coverage profile is unusual for a hybridization-based approach, which typically captures DNA extending well beyond the target region. Third, while the specificity of traditional hybridization-based methods decreases as the panel size (total kb targeted) decreases, the NEBNext Direct method retains high specificity regardless of panel size. Finally, NEBNext Direct baits for both common gene panels and custom panels are individually synthesized, then the pool of baits are screened and balanced prior to release. The balancing process adjusts the concentrations of over- and under-performing baits in the final pool so that all targets are captured within a specified window of relative efficiency. Individual bait synthesis also allows for already synthesized baits to be recombined into novel panels, and baits can be easily and quickly added or dropped from a panel as workflows undergo adjustment.

The flexibility of the NEBNext Direct protocol allows multiple sample types (such as cell-free DNA, formalin-compromised DNA, RNA, or genomic DNA), multiple sample inputs, and multiple target panels to be assayed in parallel, as each unique enrichment is treated identically from the initial hybridization to the final PCR amplification step

(at which time different cycle numbers may be required to achieve an optimal yield for all libraries). In the alternate protocols, we describe minor adjustments that should be made when working with RNA or formalin-compromised DNA inputs, all of which are performed prior to the core Basic Protocol. This all-encompassing feature, as well as careful design of each protocol step, makes NEBNext Direct unusually automation-friendly and well suited for labs with diverse target-enrichment necessities.

Strategic Planning

The following issues regarding materials and reagents should be considered prior to beginning the Basic Protocol.

Bait design

A number of decisions need to be made in order to choose the correct bait design before implementing a target-enrichment experiment or diagnostic test. First, determine if a predesigned bait set or a custom design is better suited for your goals. Predesigned panels may be found on the NEBNext Direct Web site (*https://www.neb.com/nebnext-direct/nebnext-direct-for-target-enrichment*) and are a more cost-effective option than a custom design. Currently, predesigned panels are designed for 150-bp targets and PE75 sequencing, do not include UTRs, and have an average of 10 bp of padding of the targets (extension of the target beyond the desired regions, typically into introns, to allow for optimization in bait design).

For custom designs, the desired target length and sequencing read length should be considered. For example, when targeting regions that have high homology to other regions in the genome, longer target and read lengths may be appropriate because short reads may be difficult to uniquely map to the targeted region. In contrast, if the desired targets are small, longer reads may be wasteful. With custom designs, the amount of target padding into introns should also be considered; while introns may contain variants of interest, the common occurrence of repetitive sequences within these regions may make bait design more challenging. If targets are desired in non-coding regions, common breakpoints, promoters, or other regulatory regions, specific coordinates must be defined. Finally, if designing a panel against bacterial DNA, consider whether you want to design baits for a single strain or capture and differentiate targets from multiple strains that may have minor differences in their target sequences.

Input amount

For the detection of germline variants from uncontaminated samples, only a small number of reads are required to cover a target in order to call a variant within the region with confidence. We recommend using 10 to 100 ng of genomic DNA for the detection of germline variants.

For the detection of low-frequency variants, such as somatic variants or germline variants from mixed or contaminated samples, or for the detection of mosaicism, a higher depth of coverage is typically required for confident variant calling. Therefore, we strongly recommend starting with a minimum of 100 ng of DNA input for low-frequency variant detection.

Some critical parameters, including input type, input quality, panel size, and sequencing depth, will influence the exact amount of input required for the desired coverage. As the DNA quality decreases or as the panel size increases, more input DNA is required to achieve the desired depth of coverage and subsequent sensitivity. Assuming that 10 unique reads displaying a variant allele are required to call the variant allele with confidence, the sensitivity of detection for a given input amount will be 10 divided by the target coverage (e.g., 100× coverage enables detection of a variant at 10% variant allele

Table 7.30.1 Expected Mean Coverage of Targets from a 15, 35, or 100 kb Panel Using 1000, 100, or 10 ng of High-Quality DNA as Inputs, Based on Libraries Sequenced to a Depth of 3 Million Reads by PE75 on an Illumina MiSeq

	Expected mean target coverage across panel size		
Input amount (ng)	15 kb	35 kb	100 kb
1000	6500	3000	1200
100	1500	1300	650
10	250	200	100

frequency, while 250× coverage enables variant detection at as low as 4% frequency). Refer to Table 7.30.1 to determine the DNA input amount required to achieve enough coverage for your desired sensitivity.

PCR machine

Due to the volumes of the enzymatic reactions, a PCR machine programmable to a 100 µl sample volume is necessary for the Basic Protocol. If a PCR machine is programmed for a lower sample volume, unfavorable results may be observed. Throughout the protocol, all incubations in the PCR machine (except for probe hybridization, steps 5 to 10, and library amplification, steps 72 to 77) should be performed with the PCR lid open and the heated lid off to avoid exposing the samples to high temperatures.

Capping of sample tubes

The majority of the incubations throughout the protocol, with the exceptions of probe hybridization (steps 5 to 10) and library amplification (steps 72 to 77), are at a low enough temperature that it is not necessary to cap PCR tubes to avoid evaporation. Leaving the PCR tubes uncapped may allow the protocol to be performed more quickly and make the protocol more automation friendly. However, if contamination due to open samples is a concern, the samples may be capped for each incubation and uncapped for solution exchanges.

HYBRIDIZATION-BASED METHOD TO ISOLATE TARGETS OF INTEREST FROM GENOMIC DNA

BASIC PROTOCOL

In this Basic Protocol, we describe a hybridization-based method to isolate targets of interest from genomic DNA isolated from liquid biopsies or tissue samples and convert the enriched DNA into libraries for Illumina sequencing. We additionally describe the parameters required to sequence these libraries on an Illumina MiSeq. In the subsequent Support Protocol, we suggest bioinformatic tools to analyze the sequencing data. This protocol may be easily adapted for formalin-compromised DNA, cfDNA, or RNA samples; protocols and considerations for these alternate inputs are described in the alternate protocols and Critical Parameters section below. This protocol may be used for a wide range of panel sizes, from a single gene to several hundred kilobases.

Materials

10 ng to 1 µg of input DNA
1× TE buffer, pH 8 (*APPENDIX 2A*)
Agilent High Sensitivity DNA Kit (Agilent, cat. no. 5067-4626)
NEBNext Direct Hybridization Buffer (NEB)
NEBNext Direct Hybridization Additive (NEB)
NEBNext Direct Baits (NEB)
NEBNext Direct Streptavidin Beads (NEB)

NEBNext Direct Hybridization Wash (HW; NEB)
NEBNext Direct Bead Prep Buffer (NEB)
NEBNext Direct Bead Wash 1 (BW1; NEB)
NEBNext Direct Bead Wash 2 (BW2; NEB)
NEBNext Direct 3′ Blunting Buffer (NEB)
NEBNext Direct 3′ Blunting Enzyme Mix (NEB)
NEBNext Direct dA-Tailing Buffer (NEB)
NEBNext Direct dA-Tailing Enzyme (NEB)
NEBNext Direct Adaptor Ligation Buffer (NEB)
NEBNext Direct 3′ Adaptor (NEB)
NEBNext Direct Ligase (NEB)
NEBNext Direct 5′ Blunting Buffer (NEB)
NEBNext Direct 5′ Blunting Enzyme Mix (NEB)
NEBNext Direct 5′ UMI Adaptor (NEB)
NEBNext Direct Cleaving Buffer (NEB)
NEBNext Direct Cleaving Enzyme Mix (NEB)
Molecular-biology grade water
NEBNext Direct Q5 PCR Master Mix (NEB)
NEBNext Direct Index Primers (NEB)
NEBNext Direct Sample Purification Beads (NEB)
80% (v/v) molecular-biology-grade ethanol (prepare fresh; use on same day)

Sterile, nuclease-free microcentrifuge tubes
Covaris Focused-ultrasonicator
Covaris microTubes or plate
Agilent Bioanalyzer or similar instrument
96-well PCR plates or PCR strip tubes
Eppendorf DNA LoBind 2-ml tubes (VWR, cat. no. 80077-234)
Thermocycler programmable to 100 µl
96-well plate magnet or PCR-tube magnet for working with magnetic beads
Microcentrifuge tube magnet
Illumina MiSeq sequencing system

Additional reagents and equipment for PCR (Kramer & Coen, 2000)

Fragment DNA

1. Dilute 10 ng to 1 µg of DNA to a total volume of 50 µl in 1× TE buffer, pH 8.
2. Follow the Covaris recommendations for instrument and microtube setup.
3. Shear DNA using the Covaris recommended protocol appropriate for your target size: for 150 bp targets, shear input DNA to 200 bp, and for 300 bp targets, shear input DNA to 350 bp.
4. Analyze the fragmented DNA on a Bioanalyzer or similar instrument to confirm that the desired fragment size is achieved.

Hybridize probes to DNA targets

5. Prepare a hybridization master mix for the appropriate number of samples by combining the following reagents in a sterile, nuclease-free tube. Include 10% excess if preparing the master mix for multiple samples.

   47 µl NEBNext Direct Hybridization Solution
   20 µl NEBNext Direct Hybridization Additive
   5 µl NEBNext Direct (biotinylated) Baits.

6. Mix well by vortexing or pipetting up and down 10 times.
7. Add 48 µl of fragmented DNA from steps 1 to 4 to 72 µl of hybridization master mix in a PCR tube for a final volume of 120 µl. Mix gently by pipetting up and down 10 times, then cap tubes securely to avoid evaporation.

   Throughout the protocol, samples should not be vortexed to mix unless vortexing is specifically recommended. In particular, vortexing should not be used to mix DNA-bound streptavidin beads into master mixes or wash solutions.

8. Run the following PCR program with the heated lid set to 105°C and place the samples in the thermocycler when the block temperature reaches 95°C.

   10 min, 95°C
   90 min, 60°C
   Hold, 60°C.

   Undesirable results may be obtained if the samples are started on a cold PCR block and allowed to ramp up to 95°C at the start of the PCR program.

9. While the samples are incubating, continue with the steps for preparing streptavidin beads (steps 11 to 19).
10. After the incubation at 60°C in step 8 and when preparation of streptavidin beads is complete, proceed to bead binding (step 20). Keep the samples at 60°C until ready to proceed.

Prepare streptavidin beads

11. Warm NEBNext Direct Streptavidin Beads to room temperature (~15 min).
12. Vortex the Streptavidin Beads to resuspend.
13. For each reaction, 75 µl of beads are required. Transfer the appropriate volume of beads for the number of reactions performed plus 10% excess to a sterile, nuclease-free, 2-ml microcentrifuge tube.

    To ensure thorough washing of the beads, use multiple 2-ml tubes if performing more than 12 reactions.

14. Place the tube(s) on a magnet and wait for the solution to clear (~1 min). Remove the supernatant, then remove the tube(s) from the magnet.
15. Add 150 µl of NEBNext Direct Hybridization Wash (HW) per reaction (165 µl with 10% excess) to the beads and resuspend by vortexing or pipetting up and down.
16. Place the tube(s) on a magnet and wait for the solution to clear (~1 min). Remove the supernatant, and then remove the tube(s) from the magnet.
17. Repeat steps 15 to 16 twice, for a total of three washes.
18. Resuspend the beads in 30 µl of NEBNext Direct Bead Prep Buffer per reaction (33 µl with 10% excess).
19. Keep the beads at room temperature in Bead Prep Buffer until probe hybridization (steps 5 to 10) is completed.

Bind targeted DNA to beads

20. Vortex the washed Streptavidin Beads in Bead Prep Buffer from step 18 to resuspend.

21. Uncap the samples from step 10 while keeping them in the thermocycler at 60°C. Add 30 µl of resuspended beads to each reaction, then mix gently by pipetting up and down 10 times.

    For steps 21 to 73, the thermocycler lid should be open and unheated. The PCR tubes can remain uncapped to facilitate buffer exchanges. However, if it is preferred, tubes can be capped for incubations and then decapped for the buffer exchanges.

    Take care to avoid bead loss throughout the protocol. If beads are observed on pipet tips following mixing steps, the beads should be returned to the samples before discarding the tips. Even seemingly small amounts of bead loss can result in poor library yields.

22. Change the thermocycler temperature to 48°C and incubate the reactions for 10 min.

23. Remove the samples from the thermocycler and place on a magnet. Wait for the solution to clear (~ 15 sec), remove the supernatant, and then remove the samples from the magnet.

24. Add 150 µl of HW to each sample. Gently mix by pipetting up and down 10 times. Place the samples on a thermocycler set at 62°C and incubate for 5 min.

25. Remove the samples from the thermocycler and place on a magnet. Wait for the solution to clear (~ 15 sec), remove the supernatant, and then remove the samples from the magnet.

26. Repeat steps 24 and 25 for a total of two washes at 62°C.

27. Add 150 µl of NEBNext Direct Bead Wash Buffer 2 (BW2) to each sample. Gently mix by pipetting up and down 10 times. Keep the samples at room temperature and proceed directly to step 28 to blunt 3′ ends.

    This is a safe stopping point. If you choose to stop the protocol at this point, follow Alternate Protocol 3 for overnight storage of samples mid protocol.

Blunt 3′ ends

28. While the beads are suspended in BW2 buffer, prepare a 3′ blunting master mix for the appropriate number of samples by combining the following reagents in a sterile, nuclease-free tube. Include 10% excess if preparing the master mix for multiple samples.

    97 µl NEBNext Direct 3′ Blunting Buffer
    3 µl NEBNext Direct 3′ Blunting Enzyme Mix.

29. Mix well by vortexing or pipetting up and down 10 times.

30. Place the DNA-bound beads on a magnet and wait for the solution to clear (~ 15 sec). Remove the supernatant, and then remove the samples from the magnet.

31. Add 100 µl of 3′ blunting master mix from step 28 to each sample. Gently mix by pipetting up and down 10 times.

32. Incubate the samples at 37°C for 10 min on a thermocycler with the thermocycler lid open.

33. Proceed immediately with the post-reaction wash (steps 34 to 37), then continue with step 38 to perform dA tailing.

Perform post-reaction washing

Perform all washes by pipetting up and down at least 10 times with approximately 90% of the wash volume. Insufficient mixing can result in carryover of previous reagents and

a poor library yield. Pipet the wash solutions carefully to minimize bubble formation. Some foaming of the wash solutions is to be expected.

34. Warm NEBNext Direct Bead Wash Buffer I (BW1) to room temperature prior to use.

    BW1 may contain a precipitate. If this occurs, allow the BW1 solution to reach room temperature, then gently invert the bottle ~10 times to dissolve the precipitate. Do not use BW1 until all of the precipitate has dissolved.

35. Place the samples on a magnet and wait for the solution to clear (~15 sec). Remove the supernatant, and then remove the samples from the magnet.

36. Add 150 µl of BW1 to each sample. Gently mix by pipetting up and down 10 times. Place the samples on a magnet and wait for the solution to clear (~15 sec). Remove the supernatant, then remove the samples from the magnet.

37. Add 150 µl of BW2 to each sample. Gently mix by pipetting up and down 10 times. Keep the samples at room temperature and proceed directly to the next step.

    This is a safe stopping point. If you choose to stop the protocol at this point, follow Alternate Protocol 3 for overnight storage of samples mid protocol.

Perform dA-tailing

38. While the beads are suspended in BW2 buffer, prepare a dA-tailing master mix for the appropriate number of samples by combining the following reagents in a sterile, nuclease-free tube. Include 10% excess if preparing the master mix for multiple samples.

    97 µl NEBNext Direct dA-Tailing Buffer
    3 µl NEBNext Direct dA-Tailing Enzyme.

39. Mix well by vortexing or pipetting up and down 10 times.

40. Place the DNA-bound beads on a magnet and wait for the solution to clear (~15 sec). Remove the supernatant, and then remove the samples from the magnet.

41. Add 100 µl of dA-tailing master mix from step 38 to each sample. Gently mix by pipetting up and down 10 times.

42. Incubate the samples at 37°C for 10 min on thermocycler with the thermocycler lid open.

43. Proceed immediately with the post-reaction wash (steps 34 to 37), then continue with step 44 to ligate the 3′ adaptor.

Ligate 3′ adaptor

44. While the beads are suspended in BW2 buffer, prepare a 3′ adaptor ligation master mix for the appropriate number of samples by combining the following reagents in a sterile, nuclease-free tube. Include 10% excess if preparing the master mix for multiple samples.

    80 µl NEBNext Direct Adaptor Ligation Buffer
    10 µl NEBNext Direct 3′ Adaptor
    10 µl NEBNext Direct Ligase.

45. Mix well by vortexing or pipetting up and down 10 times.

46. Place the DNA-bound beads on a magnet and wait for the solution to clear (~15 sec). Remove the supernatant, and then remove the samples from the magnet.

47. Add 100 µl of 3′ adaptor ligation master mix from step 44 to each sample. Gently mix by pipetting up and down 10 times.

48. Incubate the samples at 20°C for 15 min on a thermocycler with the thermocycler lid open.

49. Proceed immediately with the post-ligation wash (steps 50 to 53), then continue with step 54 to blunt 5′ ends.

Perform post-ligation wash

The following wash steps are different from the post-reaction washes.

50. Place the samples on a magnet and wait for the solution to clear (~15 sec). Remove the supernatant, and then remove the samples from the magnet.

51. Add 150 µl of BW1 to each sample. Gently mix by pipetting up and down 10 times. Place the samples on a magnet and wait for the solution to clear (~ 15 sec). Remove the supernatant, and then remove the samples from the magnet.

52. Repeat step 51 for a total of two washes in BW1.

53. Add 150 µl of BW2 to each sample. Gently mix by pipetting up and down 10 times. Keep the samples at room temperature and proceed directly to the next step.

    This is a safe stopping point. If you choose to stop the protocol at this point, follow Alternate Protocol 3 for overnight storage of samples mid protocol.

Blunt 5′ ends

54. While the beads are suspended in BW2 buffer, prepare a 5′ blunting master mix for the appropriate number of samples by combining the following reagents in a sterile, nuclease-free tube. Include 10% excess if preparing the master mix for multiple samples.

    97 µl NEBNext Direct 5′ Blunting Buffer
    3 µl NEBNext Direct 5′ Blunting Enzyme Mix.

55. Mix well by vortexing or pipetting up and down 10 times.

56. Place the DNA-bound beads on a magnet and wait for the solution to clear (~15 sec). Remove the supernatant, and then remove the samples from the magnet.

57. Add 100 µl of 5′ blunting master mix from step 54 to each sample. Gently mix by pipetting up and down 10 times.

58. Incubate the samples at 20°C for 10 min on a thermocycler with the thermocycler lid open.

59. Proceed immediately with the post-reaction wash (steps 34 to 37), then continue with step 60 to ligate the 5′ adaptor.

Ligate 5′ adaptor

60. While the beads are suspended in BW2 buffer, prepare a 5′ adaptor ligation master mix for the appropriate number of samples by combining the following reagents in a sterile, nuclease-free tube. Include 10% excess if preparing the master mix for multiple samples.

    80 µl NEBNext Direct Adaptor Ligation Buffer
    10 µl NEBNext Direct 5′ UMI Adaptor
    10 µl NEBNext Direct Ligase.

61. Mix well by vortexing or pipetting up and down 10 times.
62. Place the DNA-bound beads on a magnet and wait for the solution to clear (~15 sec). Remove the supernatant, and then remove the samples from the magnet.
63. Add 100 µl of 5′ adaptor ligation master mix from step 60 to each sample. Gently mix by pipetting up and down 10 times.
64. Incubate the samples at 20°C for 20 min on a thermocycler with the thermocycler lid open.
65. Proceed immediately with the post-ligation wash (steps 50 to 53), then continue with step 66 to cleave the adaptors.

Cleave adaptor

66. While the beads are suspended in BW2 buffer, prepare a cleaving master mix for the appropriate number of samples by combining the following reagents in a sterile, nuclease-free tube. Include 10% excess if preparing the master mix for multiple samples.

    95 µl NEBNext Direct Cleaving Buffer
    5 µl NEBNext Direct Cleaving Enzyme Mix.

67. Mix well by vortexing or pipetting up and down 10 times.
68. Place the DNA-bound beads on a magnet and wait for the solution to clear (~15 sec). Remove the supernatant, and then remove the samples from the magnet.
69. Add 100 µl of cleaving master mix from step 66 to each sample. Gently mix by pipetting up and down 10 times.
70. Incubate the samples at 37°C for 15 min on a thermocycler with the thermocycler lid open.
71. Proceed immediately with the post-reaction wash (see steps 34 to 37), then continue to the step for library amplification (step 72).

Perform library amplification

Essential protocols for PCR are described in Kramer & Coen (2000).

72. Place the reactions on a magnet, wait for the solution to clear (~15 sec), remove the supernatant, then remove the reactions from the magnet.
73. Add 45 µl of molecular-biology-grade water to each reaction. Mix gently by pipetting up and down 10 times to completely resuspend the beads.
74. Combine the following components in a new, sterile, nuclease-free PCR tube:

    50 µl NEBNext Direct Q5 PCR Master Mix
    5 µl NEBNext Direct Index Primer Mix
    45 µl resuspended beads from step 73.

    Choose appropriate indexed primers following the NEBNext Direct guidelines for pooling multiple barcoded samples (https://www.neb.com/nebnext-direct/nebnext-direct-for-target-enrichment).

75. Gently mix by pipetting up and down 10 times. Securely cap the PCR tubes.
76. Run the following program with the heated lid set to 105°C and place the samples in the thermocycler when the block temperature reaches 98°C:

1 cycle:	30 sec	98ºC	(initial denaturation)
15 to 25 cycles:	10 sec	98ºC	(denaturation)
	15 sec	62ºC	(annealing)
	20 sec	72ºC	(extension)
1 cycle:	5 min	72ºC	(final extension)
1 cycle:	indefinitely	4ºC	(hold).

The number of PCR cycles required will vary depending on the amount of input material. Follow the guidelines below to determine the appropriate number of cycles:

1000 ng input DNA, 20 cycles
500 ng input DNA, 21 cycles
100 ng input DNA, 23 cycles
10 ng input DNA, 25 cycles.

Undesirable results may be observed if the samples are started on a cold PCR block and are allowed to ramp up to 98°C at the start of the PCR program.

77. Proceed to purify and size select amplified fragments.

PCR reactions with beads can be stored for up to 16 hr at 4°C.

Purify and size select amplified fragments

If you detect significant evaporation from the PCR reaction, bring the volume up to 100 µl with molecular-biology-grade water.

78. Warm NEBNext Direct Sample Purification Beads to room temperature.
79. Vortex the Sample Purification Beads to resuspend.
80. Add 85 µl of Sample Purification Beads to each PCR reaction. Mix well by pipetting up and down at least 10 times.
81. Incubate the samples uncapped for 10 min at room temperature.
82. Place the samples on a magnet. After the solution is clear (about 2 min), carefully remove and discard the supernatant.

During steps 80 to 85, the beads contain the DNA targets. Be careful not to disturb or discard the beads during these steps.

83. Add 200 µl of freshly prepared (same day) 80% ethanol while the samples are on the magnet. Incubate at room temperature for 30 sec, then carefully remove and discard the supernatant.
84. Repeat step 83 once for a total of two washes in 80% ethanol, ensuring that all of the supernatant is removed from each reaction.
85. Incubate the samples uncapped at 37°C for 5 min on a thermocycler with the thermocycler lid open to dry the beads.
86. Remove the tubes from the thermocycler and resuspend the dry beads in 102 µl of water. Incubate for 2 min at room temperature.

During this step, the DNA targets are eluted into the supernatant and remain in the supernatant though step 88. Be careful not to discard the supernatant during these steps.

87. Place the tubes on a magnet and allow the solution to clear (about 2 min).

88. Transfer 100 µl of the eluted library to fresh tubes and add 85 µl of Sample Purification Beads. Mix well by pipetting up and down at least 10 times.

89. Incubate for 10 min at room temperature.

90. Place the tubes on a magnet. After the solution is clear (about 2 min), carefully remove and discard the supernatant.

    During steps 90 to 93, the beads contain the DNA targets. Be careful not to disturb or discard the beads during these steps.

91. Add 200 µl of freshly prepared (same day) 80% ethanol while the samples are on the magnet. Incubate at room temperature for 30 sec, then carefully remove and discard the supernatant.

92. Repeat step 91 once for a total of two washes in 80% ethanol, ensuring that all of the supernatant is removed from each well.

93. Incubate the samples uncapped at 37°C for 2 min on a thermocycler with the thermocycler lid open to dry the beads.

94. Remove the tubes from the thermocycler and resuspend the dry beads in 30 µl of 1× TE buffer, pH 8, by gently pipetting (or gently vortex capped tubes, then quickly spin down). Incubate for 2 min at room temperature.

    During this step, the DNA targets are eluted into the supernatant. Be careful not to discard the supernatant.

95. Place the tubes on a magnet and allow the solution to clear (about 2 min).

96. Transfer 28 µl of the eluted library to a fresh tube.

97. Evaluate the size distribution of the eluted library using an Agilent High Sensitivity Bioanalyzer Chip or similar approach.

98. Accurately quantify your library using a method discussed in Critical Parameters, below.

Run samples on the Illumina MiSeq

The NEBNext Direct protocol incorporates Illumina adaptor sequences; therefore, the libraries generated from this protocol may be sequenced on any Illumina platform. Here we describe the steps necessary to sequence NEBNext Direct libraries on the Illumina MiSeq.

99. Follow the Illumina MiSeq Reporter Software Guide to reconfigure the MiSeq reporter parameter CreateFastqForIndexReads to write both index reads to FASTQs.

    The sample barcode is in the i7 position and is sequenced as the index read 1 (I1). The UMI is in the i5 position and is sequenced as the index read 2 (I2). By default, the Illumina MiSeq is set not to generate FASTQs for index reads; therefore, it is necessary to override this setting in order to make use of these features.

    If samples are run on the Illumina NextSeq, please note that while the i5 index is generated in the reverse complement orientation, no adjustments need to be made because the i5 UMI is a random sequence.

100. Generate a MiSeq sample sheet with the following parameters:

    a. Workflow: Generate FASTQ.
    b. Application: FASTQ only.
    c. Assay: TruSeq HT.
    d. Chemistry: Amplicon.

The remaining fields will require information specific to your experiment. The choice for read lengths will depend on the particular bait design for your experiment, but in general NEBNext Direct baits are designed for 2 × 75PE reads. For a sample sheet template, visit the NEBNext Direct Web site at:

https://www.neb.com/nebnext-direct/nebnext-direct-for-target-enrichment.

101. Pool, dilute, and denature samples for an 8 pM final concentration following the MiSeq Denature and Dilute Libraries Guide and the NEBNext Direct guidelines for pooling together multiple barcoded samples (*https://www.neb.com/nebnext-direct/nebnext-direct-for-target-enrichment).*

    The number of samples that can be pooled together will depend on the input amount, panel size, and number of reads required from each sample for the particular analysis being performed. While the total number of reads acquired from a MiSeq run will vary depending on the cluster density, type of flowcell, and version of the MiSeq reagent kit, Table 7.30.1 may be used as a starting point to determine pooling conditions.

102. Follow the MiSeq System Guide to load the samples and run the MiSeq.

SUPPORT PROTOCOL 1

ANALYSIS OF SEQUENCING READS

In these steps, sequencing reads are aligned, library metrics are generated, and genetic variants are called. There are many publicly available tools to accomplish these steps, including software that is available for free and typically used through a command-line interface, as well as more user-friendly options that can be purchased from analysis software providers. Here we describe one approach that may be used to analyze sequencing data using the field-standard, publically available Picard tools (see *http://broadinstitute.github.io/picard*).

1. Convert FASTQ reads into an unmapped bam file using Picard's FastqToSam tool.
2. Identify and mark Illumina adaptor sequences with Picard's MarkIlluminaAdapters tool and the default Illumina adaptor sets.
3. Convert files with marked adaptor sequences into interleaved FASTQs using Picard's SamToFastq. An interleaved FASTQ is a single file that contains both read 1 and read 2.
4. Align the interleaved FASTQ file to the appropriate genome using BWA-MEM (Li & Durbin, 2010), available for download at *http://bio-bwa.sourceforge.net/*.
5. Merge separate alignments using Picard's MergeBamAlignment to create a final BAM file.
6. Mark duplicates by considering both the start and stop coordinates of the alignments as well as the UMIs (see *https://github.com/fulcrumgenomics/fgbio*). To work with the UMIs, use Picard's MarkDuplicates with the option `BARCODE_TAG=RX` (or any alternative tag that was selected). Flag duplicates in the BAM file but do not remove them. This BAM file will be used to generate various metrics with Picard tools.
7. Generate quality control information using the following Picard tools: AlignmentSummaryMetrics, InsertSizeMetrics, QualityScoreDistribution, QualityByCycle, and BaseDistributionMetrics.
8. Evaluate off target reads by finding regions outside of the targets that have coverage $\geq 10\times$.
9. To produce further analysis metrics, run Picard's HsMetrics on both pre- and post-duplicate-marked reads.

10. To perform variant calling, filter BAM files to only include high-quality reads (we use a custom script to remove reads with a mapping quality <10, pairs in which one read is unmapped, duplicates, and secondary alignments).

11. Use the Genome Analysis ToolKit (GATK) tools RealignerTargetCreator and IndelRealigner to identify putative insertions and deletions (indels) and to realign reads to these regions.

    See https://software.broadinstitute.org/gatk/gatkdocs/ for more details on GATK tools, including the databases used to identify known indels.

12. Use the GATK tools Haplotypecaller, GenotypeGVCFs, and Picard's FilterVcf to generate a GVCF file for each sample, perform joint genotype calling for all samples, and filter the files for high-quality variants.

    Some variant callers use target region lists in BED file format. Use Picard's IntervalListToBed to convert files in intervalList format to bed files. Use Picard's IntervalListTools for optional commands to manipulate target files, including padding targets to detect nearby variants.

13. A variety of tools may be used to detect somatic variants, including Scalpel (Narzisi et al., 2014), Mutect (Cibulskis et al., 2013), and Vardict (Lai et al., 2016). Tumor-normal pairs or single samples may be used as inputs. The resulting vcf files can be combined, filtered, and annotated to indicate known variants from databases such as dbSNP (Sherry et al., 2001) or COSMIC (Forbes et al., 2015).

    For more details and command-line examples see the NEBNext Direct FAQ page: https://www.neb.com/~/media/NebUs/Files/PDF%20FAQ/NEBNext%20Direct%20Cancer%20HotSpot%20Panel%20FAQ_4-27.pdf.

NEBNext DIRECT TARGET ENRICHMENT FOR FORMALIN-COMPROMISED DNA

ALTERNATE PROTOCOL 1

Tumor samples from patients are often fixed in formalin and stored as formalin-fixed, paraffin-embedded (FFPE) blocks with the intention to preserve the morphology of the tumor sample for microscopy. However, FFPE samples are often a necessary and valuable source of DNA in cancer studies, especially when archived samples from biobanks are used. Formalin fixation can compromise the quality of DNA and cause degradation and chemical modifications (Srinivasan, Sedmak, & Jewell, 2002), making FFPE samples a challenging source of DNA for NGS applications (Hedegaard et al., 2014). While the NEBNext Direct technology tolerates internally damaged DNA, damage to the DNA ends is incompatible with adaptor ligation. Thus, we recommend following the protocol below when working with FFPE DNA to repair the 3′ and 5′ ends prior to target enrichment and library preparation.

Additional Materials *(also see Basic Protocol 1)*

10 ng to 1 µg formalin-compromised DNA
NEBNext Direct FFPE Phosphorylation Enzyme (NEB)
NEBNext Direct FFPE Phosphorylation Buffer (NEB)

1. Shear fragmented DNA according to steps 1 to 4 of the Basic Protocol.

2. Prepare an FFPE phosphorylation master mix for the appropriate number of samples by combining the following reagents in a sterile, nuclease-free tube. Include 10% excess if preparing the master mix for multiple samples.

   5 µl NEBNext Direct FFPE Phosphorylation Buffer
   1 µl NEBNext Direct FFPE Phosphorylation Enzyme.

3. Mix 42 µl of the fragmented FFPE DNA from step 1 with 6 µl of FFPE phosphorylation master mix from step 2 in a sterile, nuclease-free tube.
4. Mix gently by pipetting up and down five times.
5. Incubate at 37°C for 15 min with the heated lid set to 45°C
6. Briefly spin down the reaction to collect the sample to the bottom of the tube.
7. Place the samples on ice and proceed directly to probe hybridization (step 5 of the Basic Protocol), continuing with the subsequent steps.

ALTERNATE PROTOCOL 2

NEBNext DIRECT TARGET ENRICHMENT FOR RNA SAMPLES

RNA-seq is a powerful tool that enables the analysis of gene expression as well as the detection of gene fusions and transcriptional isoforms. By combining target enrichment with RNA-seq, coverage of transcripts of interest is increased at decreased sequencing costs. To perform target enrichment for RNA-seq using the NEBNext Direct approach, minor adjustments need to be made to use the RNA as input for the Basic Protocol. In this alternate protocol, total RNA (20 ng to 1 µg) is converted into double-stranded cDNA (dscDNA) using the NEBNext® RNA First and Second Strand Synthesis Modules (NEB, cat. nos. E7525S/L and E6111S/L) prior to target enrichment and library preparation. The generation of dscDNA enables the capture of gene fusions when only one fusion partner is known, regardless of whether the known gene is 3′ or 5′ of the fusion site. This protocol is optimized for RNA inserts of approximately 200 bp and should be sequenced with 2 × 150PE reads.

Additional Materials *(also see Basic Protocol)*

20 ng to 1 µg total RNA
5× NEBNext First Strand Synthesis Reaction Buffer
NEBNext Random Primers
Murine RNase inhibitor (NEB)
ProtoScript II Reverse Transcriptase (NEB)
10× Second-Strand Synthesis Reaction Buffer (NEB)
Second-Strand Synthesis Enzyme Mix (NEB)

RNA fragmentation and first strand synthesis

1. Dilute the total RNA (20 ng to 1 µg) to a total volume of 5 µl with sterile, nuclease-free water in a sterile, nuclease-free PCR tube and keep on ice.
2. Prepare a fragmentation and priming master mix for the appropriate number of samples by combining the following reagents in a sterile, nuclease-free tube on ice. Include 10% excess if preparing the master mix for multiple samples.

   4 µl 5× NEBNext First Strand Synthesis Reaction Buffer
   1 µl NEBNext Random Primers.
3. On ice, add 5 µl of fragmentation and priming master mix to the PCR tube containing 5 µl total RNA for a final volume of 10 µl. Mix gently by pipetting up and down and briefly microcentrifuge the tube to bring down the contents.
4. Incubate the sample in a preheated thermocycler set to 94°C for 15 min with the heated lid set to 105°C, and then transfer the tube to ice.
5. Combine the following reagents for first-strand synthesis in a sterile, nuclease-free tube on ice. Include 10% excess if preparing the enzyme mix for multiple samples.

   0.5 µl murine RNase inhibitor

1 µl ProtoScript II Reverse Transcriptase
8.5 µl nuclease-free H_2O.

6. On ice, add 10 µl of first-strand synthesis enzyme mix to the PCR tube containing 10 µl of fragmented and primed RNA from step 4 for a final volume of 20 µl. Mix gently by pipetting up and down, briefly spin the tube to bring down the contents, and keep on ice.

7. Run the following PCR program with the heated lid set to 105°C and place the samples in the thermocycler when the block temperature reaches 25°C.

   10 min, 25°C
   50 min, 42°C
   15 min, 70°C
   Hold, 4°C.

8. Place the heat-inactivated cDNA reaction on ice. Add 48 µl of nuclease-free water to the reaction. Proceed directly with second-strand synthesis.

Second-strand synthesis

9. Combine the following reagents for second-strand synthesis in a sterile, nuclease-free tube on ice. Include 10% excess if preparing the enzyme mix for multiple samples.

   8 µl 10× Second Strand Synthesis Reaction Buffer
   4 µl Second Strand Synthesis Enzyme Mix.

10. On ice, add 12 µl of the second-strand synthesis master mix to the PCR tube containing 68 µl of diluted first-strand cDNA for a final volume of 80 µl. Mix gently by pipetting up and down, briefly spin the tube to bring down the contents, and keep on ice.

11. Incubate in a thermal cycler for 1 hr at 16°C. Proceed immediately with sample purification.

Sample purification

12. Warm Sample Purification Beads to room temperature and vortex to resuspend the beads.

13. Add 144 µl (1.8×) of resuspended Sample Purification Beads to the second-strand synthesis reaction and mix well by pipetting up and down at least 10 times.

14. Incubate for 5 min at room temperature.

15. Place the tube on a magnet to separate the beads from the supernatant. After the solution is clear (about ~15 sec), carefully remove and discard the supernatant.

    During steps 13 to 18, the beads contain the dscDNA targets. Be careful not to disturb or discard the beads during these steps.

16. Add 200 µl of freshly prepared (same day) 80% ethanol while the samples are on the magnet. Incubate at room temperature for 30 sec, then carefully remove and discard the supernatant.

17. Repeat step 16 once for a total of two washes in 80% ethanol, ensuring that all of the supernatant is removed from each reaction.

18. Incubate the samples uncapped at 37°C for 5 min on a thermocycler with the thermocycler lid open to dry the beads.

19. Remove the tubes from the thermocycler and resuspend the dry beads in 50 µl of water. Incubate for 2 min at room temperature.

    During this step, the dscDNA targets are eluted into the supernatant and remain in the supernatant. Be careful not to discard the supernatant.

20. Carefully transfer 48 µl of the dscDNA into a fresh PCR tube or plate and proceed to step 4 of the Basic Protocol, continuing with the subsequent steps.

    Purified dscDNA samples can be stored overnight at 4°C or at −20°C for long-term storage.

ALTERNATE PROTOCOL 3

OVERNIGHT STORAGE OF SAMPLES MID-PROTOCOL

Throughout the Basic Protocol, there are several safe stopping points at which samples may be stored in 1 × TE buffer for up to 16 hr at 4°C. These stopping points are indicated in the Basic Protocol.

For materials, see Basic Protocol.

1. Place the samples on the magnet and allow the solution to clear (~15 sec).
2. Remove the supernatant, then remove the samples from the magnet.
3. Resuspend the beads in 100 µl of 1 × TE buffer, pH 8.
4. Store the DNA-bound beads for up to 16 hr at 4°C.
5. When ready to proceed with the protocol, place the samples on the magnet and allow the solution to clear (~15 sec).
6. Remove the supernatant, then remove the samples from the magnet.
7. Add 150 µl of BW2 to each reaction and mix gently by pipetting up and down 10 times.
8. Proceed directly with the next step in the protocol.

COMMENTARY

Background Information

There are a variety of techniques that may be used to enrich specific genomic loci prior to sequencing. Each approach has advantages and disadvantages that must be considered for any given application. Several parameters should be considered when choosing the appropriate approach, such as the target territory size, amount of input material, protocol length, and data type desired. Many of these techniques are commercially available and fall into one of two major categories—in-solution hybridization or multiplex PCR—while several newer technologies incorporate features of both approaches.

In-solution hybridization approaches

In-solution hybridization-based approaches, originally developed for whole-exome sequencing (Gnirke et al., 2009), use biotinylated oligonucleotides to capture genomic regions of interest. Commercially available kits use DNA or RNA baits ranging from 50 to 150 nucleotides. Specificity for targeted regions is typically high for larger panels and decreases as the size of the targeted region decreases. Thus, small panels typically result in an increased proportion of sequencing that is lost to off-target regions.

Many commercially available products use microarrays to synthesize oligonucleotides, followed by PCR to amplify the pool of baits. While array-based synthesis produces large sets of baits more efficiently than individual synthesis, this process makes it difficult to adjust bait concentrations for even coverage of targets or for removal of baits that are no longer desired.

Many in-solution hybridization protocols have long and complex workflows. Target enrichment generally follows library generation from the whole genome, and begins with complicated and long hybridization and washing steps. The typical length of an in-solution protocol ranges from 18 to 36 hr. Because amplification of the libraries precedes hybridization, PCR copies of target molecules are difficult to identify and discriminate from unique molecules in the starting material, and uneven amplification across targets makes determining true allelic frequencies of variants difficult. Finally, the initial library generation

step requires dsDNA; thus, the approach will lose any single-stranded DNA that may be present in the original sample due to DNA damage.

Multiplex PCR approaches

Multiplex PCR-based enrichment typically targets a smaller territory than in-solution hybridization and is typically limited to 150 to 200 amplicons (Mertes et al., 2011). Using a pool of primers, enrichment is accomplished through PCR amplification of the targeted regions. Multiplex PCR-based workflows are relatively fast, require low amounts of input material, and produce high specificity for targeted regions. However, primer design is a major challenge because melting temperatures must match within each panel and primer-primer interactions must be considered. These constraints can lead to variations in coverage uniformity between targets. Partitioning individual amplification reactions into emulsion droplets can alleviate some of these constraints and improve target uniformity (Tewhey et al., 2009), but this approach requires investment in a digital droplet PCR machine. Multiplex PCR also presents challenges in targeting samples that are degraded, as the probability of landing both primers on a given molecule decreases with shorter fragments of input DNA molecules.

Hybrid technologies

Because hybridization-based panels lose specificity as they decrease in size and PCR-based panels become more difficult to design and keep uniform as they increase in size, no single hybridization- or PCR-based technology can be used across a range of applications. Additionally, panels that fall in a middle range with respect to size are challenged by the shortcomings of both approaches. Thus, newer technologies attempt to bridge the gap between hybridization and PCR-based approaches. Examples of these hybrid approaches include multiplex extension ligation (Shen et al., 2011), molecular inversion probes (MIPS)/padlock probes (Porreca et al., 2007), nested patch PCR (Varley & Mitra, 2008), and selector probes (Johansson et al., 2011), and, more recently, the NEBNext Direct approach for target enrichment.

The NEBNext Direct approach eliminates many of the drawbacks that exist in other enrichment technologies. Unlike other in-solution hybridization protocols, the NEBNext Direct protocol has no library preparation prior to hybridization. This feature reduces the amount of amplification that is required throughout the protocol and enables single-stranded DNA to be used for library construction. Additionally, the NEBNext Direct approach allows for filtering of PCR duplicates prior to variant calling, thus removing false positive mutations that can arise during PCR. Unique molecules are identified by two features of the NEBNext Direct library: a variable 5′ end and a 12-bp randomized unique molecular identifier that is incorporated into the 5′ adaptor. Finally, the NEBNext Direct approach maintains high specificity across a broad range of target territory, from single genes or exons to hundreds of kilobases, eliminating the need to use different technologies for different panels.

The target coverage profiles of NEBNext Direct libraries are unique for a hybridization-based approach in that reads have a defined 3′ end, resulting in coverage profiles that resemble those of PCR-based libraries (Fig. 7.30.2). This differs from a typical hybridization-based approach, in which randomly fragmented molecules are captured and read coverage resembles a normal distribution. The defined 3′ end of the NEBNext Direct libraries is produced by truncating the capture molecule up to the bait/DNA complex, and results in less off-target sequencing of regions near the targeted loci. The coverage profiles of NEBNext Direct libraries additionally demonstrate even coverage across the targeted regions, which is a result of individually synthesized baits that are balanced to reduce the concentrations of over-performing baits.

The NEBNext Direct approach is highly amenable to laboratory automation. Unlike alternative hybridization-based approaches, the technique does not require hybridization buffers or post-capture washes to be preheated, which introduces complexity into the workflow and can lead to variable results. Aside from the hybridization and library amplification steps, sample incubations occur at temperatures between 20º and 62°C, allowing for the samples to remain uncapped throughout the majority of the protocol. Additionally, captured fragments remain bound to magnetic streptavidin beads throughout the protocol, which minimizes sample transfers between vessels and simplifies buffer exchanges.

The NEBNext Direct protocol has some limitations that should be considered prior to beginning an experiment. First, NEBNext Direct relies on mechanical fragmentation, which generates highly reproducible fragmentation

of genomic DNA but requires special equipment and may result in some sample loss due to material transfers in and out of customized labware. Second, the post-enrichment conversion of captured fragments into sequencing-ready libraries entails a series of enzymatic steps with interspersed bead washes that require careful pipetting to prevent the loss of material throughout the protocol. Third, while an input range of 10 ng to 1 µg is sufficient for germline variant detection, 100 ng of material is recommended to call somatic variants, which may present challenges if limited amounts of material are available. Finally, the protocol involves redundant pipetting, which presents challenges when performed manually but is ideal for automation.

Critical Parameters

Input type

A variety of input types are compatible with this protocol. The Basic Protocol outlined above is the standard protocol for genomic DNA. Examples of DNA sources that are compatible with the Basic Protocol include blood, fresh-frozen tissue, cell cultures, cheek swabs, or saliva.

When working with FFPE DNA, begin the procedure using Alternate Protocol 1 (NEBNext Direct target enrichment for formalin-compromised DNA), then continue with the Basic Protocol. Depending on its quality, FFPE DNA may perform as well as genomic DNA, or if it is highly damaged, a lower conversion rate may be observed. It is important to properly quantify amplifiable FFPE DNA before performing target enrichment and library construction; when highly damaged FFPE DNA is used, larger inputs may be required.

Cell-free DNA (cfDNA) isolated from blood is already fragmented to an ideal size for this protocol; therefore, steps 1 to 4 of the protocol should be skipped and the cfDNA should be used directly for probe hybridization. We typically see the highest conversion rate for cfDNA due to its size and optimal 3′ and 5′ ends.

If working with RNA samples, the RNA must first be converted to dscDNA before beginning the Basic Protocol. Follow Alternate Protocol 2 (NEBNext Direct target enrichment for RNA samples) to convert the samples to dscDNA, then use the dscDNA as the input for probe hybridization. The integrity of the RNA sample will impact the conversion rate of the sample. We recommend analyzing the RNA sample on a Bioanalyzer to assess the quality of the RNA; a minimum RNA Integrity Number (RIN) of 8 is recommended for this protocol. If working with RNA samples, the enrichment of desired targets over ribosomal RNA will vary depending on the expression levels of the desired targets; however, with highly expressed transcripts, we typically observe a value greater than 90% for the Picard metric `PCT_SELECTED_BASES`. When attempting to enrich a target with low expression levels, ribosomal depletion (such as the NEBNext rRNA Depletion Kit) or mRNA selection (such at the NEBNext Poly(A) mRNA Magnetic Isolation Module) may be necessary.

Input quantification

Accurate quantification of input DNA or RNA is essential for successful library preparation. Input samples should be quantified using a sensitive method such as a High Sensitivity Qubit Assay. Formalin-damaged samples should be quantified using an approach that measures the number of amplifiable molecules rather than total nucleic acid content, such as the Agilent NGS FFPE QC Kit (Agilent Technologies).

Panel size

This protocol accommodates a wide range of panel sizes, from a single gene to several hundred kilobases. The same Basic Protocol may be used across panel sizes without modification to the protocol. The size of the panel will influence the yield of the library; smaller panels will produce a lower yield overall but will have greater unique coverage of the targeted area, while larger panels will have a higher overall yield, but coverage per target will be lower.

Number of samples

We recommend using a multichannel pipettor when working with several samples and using a PCR strip tube or PCR plate to facilitate handling of the samples. If working with multiple samples, master mixes should be prepared with 10% excess. This protocol may be easily automated on most liquid handlers to accommodate high volumes of samples.

Library quantification

Accurate library quantification is essential to properly load the sequencer for the desired depth of sequencing. Therefore, it is important to quantify the final library product using

Table 7.30.2 Troubleshooting

Library appearance	Potential cause	Resolution
No library, no adaptor dimer	1. Forgot to add PCR primers 2. Discarded library during sample purification bead cleanup	Repeat library construction
Abundant adaptor dimer with acceptable library yield	Non-protocol bead cleanup method	Repeat bead cleanup with a 0.85:1 bead:sample ratio
Abundant adaptor dimer with poor library yield	1. Insufficient post-reaction or post-ligation washing with BW1 and BW2 2. Streptavidin bead loss	1. Repeat library construction; pipet up and down 10 times with 90% of the wash volume for each BW1 and BW2 wash 2. Repeat library construction; check pipet tips carefully following each wash or mix step to ensure no beads remain in pipet tips
Poor library yield with FFPE-treated DNA	DNA is highly damaged	Repeat library construction and follow Alternate Protocol 1 to end-repair FFPE DNA. Use more input DNA and accurately quantify amplifiable DNA molecules using a qPCR based method.
Poor library yield with non-FFPE DNA	1. DNA was not accurately quantified 2. Too few amplification cycles for the amount of input DNA	1. Re-quantify input DNA with a more reliable method and use a minimum of 10 ng 2. Use number of PCR cycles recommended in the guidelines in step 76 of the Basic Protocol
Large library yield with abundant high-molecular-weight DNA	Too many amplification cycles for the amount of input DNA or panel size	Use the number of PCR cycles recommended in the guidelines in step 76 of the Basic Protocol

a highly sensitive approach. There are several options for quantifying the sequencing-ready library. At a minimum, the final libraries should be analyzed and quantified on a Bioanalyzer or similar instrument. For a more sensitive quantification assay, a qPCR approach such as the qPCR NGS Library Quantification Kit from Agilent Technologies, the KAPA library quantification kits for Illumina platforms, or the NEBNext Library Quant Kit for Illumina may be used.

Troubleshooting

The troubleshooting guidance in Table 7.30.2 does not account for failing to add baits or target DNA to the hybridization step or for skipping protocol steps or performing them incorrectly. These errors will typically result in failure to generate a library.

Anticipated Results

The results of NEBNext Direct target enrichment are determined at two separate time points: (1) directly after library construction, and (2) after sequencing.

The quality of library construction is assayed using an Agilent Bioanalyzer High Sensitivity DNA chip. Library appearance and yield depend on input DNA amount, input DNA type, and the number of individual baits in the hybridization reaction (Fig. 7.30.3). In general, library shape should resemble the size distribution of input DNA; sonicated DNA

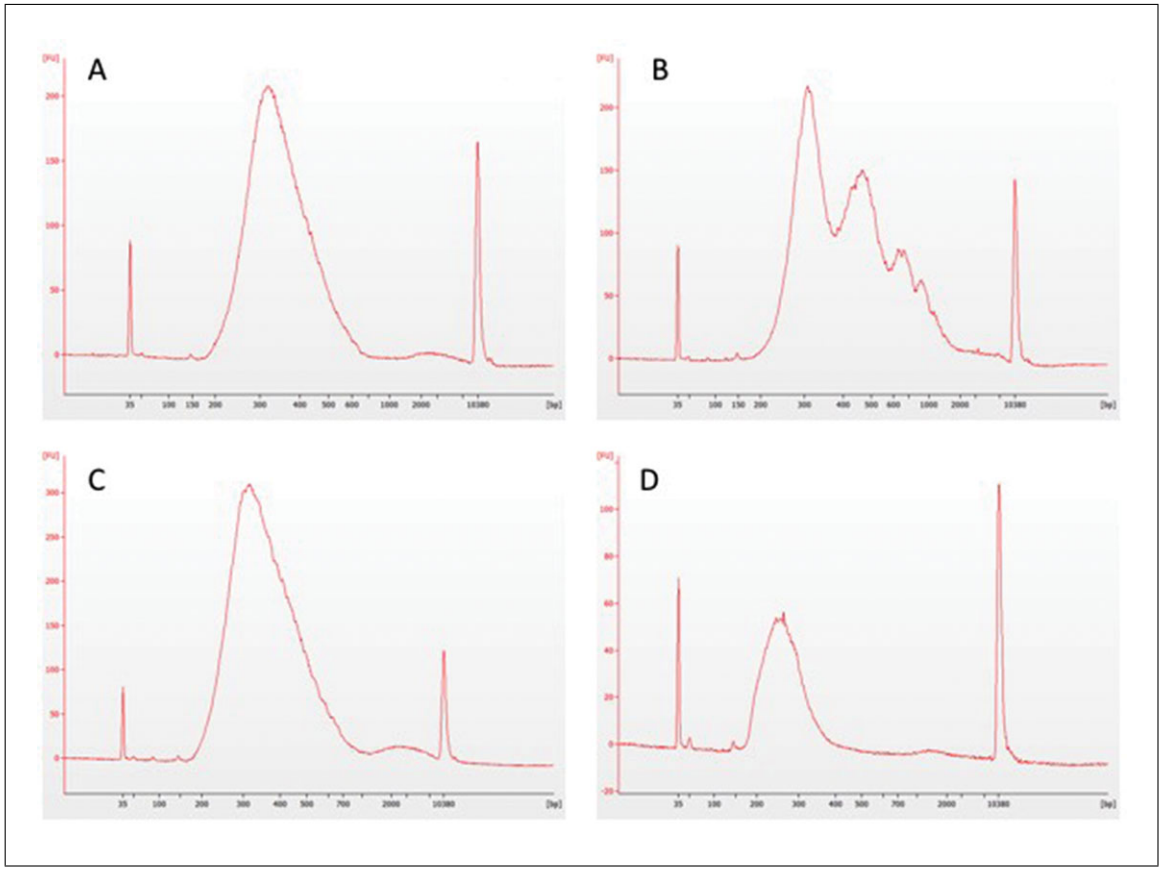

Figure 7.30.3 Representative Bioanalyzer traces of libraries prepared from different input types using the NEBNext Direct protocol. (**A**) 100 ng of sonicated DNA. (**B**) 25 ng of cfDNA. (**C**) 100 ng of FFPE DNA from colon. (**D**) 100 ng of FFPE DNA from liver.

produces a single peak (Fig. 7.30.3A) while cfDNA produces a stereotypical, multi-peak library (Fig. 7.30.3B). A major factor that will affect yield is DNA quality. FFPE DNA is chemically damaged to varying degrees, and library yields are correspondingly variable (see libraries in Figure 7.30.3C and 7.30.3D, both produced from 100 ng FFPE DNA from different sources). Phosphorylation treatment (see Alternate Protocol 1) can help to repair damaged DNA ends and boost yield for FFPE libraries. In contrast, cfDNA is naturally fragmented in the body by nucleases (Snyder, Kircher, Hill, Daza, & Shendure, 2016), which produce 3′ and 5′ ends that, when used with the single-stranded DNA capture and 3′ and 5′ blunting features of the NEBNext Direct approach, are optimal for adaptor ligation. Consequently, NEBNext Direct libraries produced from cfDNA are typically of a higher yield than would be expected for the equivalent amount of sonicated DNA. Finally, all other factors being equal, a larger bait set produces a larger yield.

After sequencing, data analysis with standard Picard tools returns several metrics that reflect the quality of the target enrichment (see Table 7.30.3). Some of these metrics vary depending on input type. In particular, FFPE DNA frequently shows lower alignment and on-target rates. Depth of target coverage is directly related to the amount of input DNA used and the number of reads generated in the sequencing run; more input DNA and more reads results in greater depth of coverage.

Time Considerations

The conversion of fragmented DNA to a sequencing-ready library using the NEBNext Direct Basic Protocol (from probe hybridization through purification and size selection of amplified products) can be completed in less than 7 hr (approximately 2.5 hr for probe hybridization through bead binding, 2 hr for 3′ blunting through adaptor cleavage, 1 hr to set up and run the library amplification, and 1 hr for purification and size selection of amplified products). If necessary, there are several convenient stopping points throughout the protocol to divide the procedure over more than 1 day. Additional time is required to analyze and quantify the completed library, to run the

Table 7.30.3 Sample Metrics from the NEBNext Direct Cancer Hot Spot Panel, which Covers a 37-kb Targeted Area

% Reads Aligned	99.2%
% Reads Paired	100%
% Reads On Target	93.6%
% Bases On Target	61.6%
Normalized Coverage of Targets with GC 0-35%	0.8
Normalized Coverage of Targets with GC 35-65%	1.0
Normalized Coverage of Targets with GC 65-100%	0.7
Median Insert Size	174
Fold Enrichment	51,367
% Bases > 50% Mean	92.9%
% Bases > 33% Mean	98.1%
% Bases > 25% Mean	99.5%
% Bases >20% Mean	100%
Fold 80 Base Penalty	1.42

sequencer, and to analyze data; the time considerations for these steps will vary depending on the methods of choice.

Acknowledgements

The authors would like to acknowledge Tim Fennell of Fulcrum Genomics for providing the UMI duplicate marking solution and additional data analysis support and Isabel Gautreau of New England Biolabs for her input with troubleshooting.

Literarature Cited

Cibulskis, K., Lawrence, M. S., Carter, S. L., Sivachenko, A., Jaffe, D., Sougnez, C., ... Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnology*, *31*, 213–219. doi: 10.1038/nbt.2514.

Forbes, S. A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., ... Campbell, P. J. (2015). COSMIC: Exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Research*, *43*, D805–811. doi: 10.1093/nar/gku1075.

Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E. M., Brockman, W., ... Nusbaum, C. (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology*, *27*, 182–189. doi: 10.1038/nbt.1523.

Hedegaard, J., Thorsen, K., Lund, M. K., Hein, A.-M. K., Hamilton-Dutoit, S. J., Vang, S., ... Dyrskjøt, L. (2014). Next-generation sequencing of RNA and DNA isolated from paired fresh-frozen and formalin-fixed paraffin-embedded samples of human cancer and normal tissue. *PLOS One*, *9*, e98187. doi: 10.1371/journal.pone.0098187.

Johansson, H., Isaksson, M., Sörqvist, E. F., Roos, F., Stenberg, J., Sjöblom, T., ... Nilsson, M. (2011). Targeted resequencing of candidate genes using selector probes. *Nucleic Acids Research*, *39*, e8. doi: 10.1093/nar/gkq1005.

Kozarewa, I., Armisen, J., Gardner, A. F., Slatko, B. E., & Hendrickson, C. L. (2015). Overview of target enrichment strategies. *Current Protocols in Molecular Biology*, *112*, 7.21.1–7.21.23. doi: 10.1002/0471142727.mb0721s112.

Kramer, M. F., & Coen, D. M. 2001. Enzymatic amplification of DNA by PCR: Standard procedures and optimization. *Current Protocols in Molecular Biology*, *56*, 15.1:15.1.1–15.1.14. doi: 10.1002/0471142727.mb1501s56.

Lai, Z., Markovets, A., Ahdesmaki, M., Chapman, B., Hofmann, O., McEwen, R., ... Dry, J. R. (2016). VarDict: A novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Research*, *44*, e108. doi: 10.1093/nar/gkw227.

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, *26*, 589–595. doi: 10.1093/bioinformatics/btp698.

Mertes, F., ElSharawy, A., Sauer, S., van Helvoort, J. M. L. M., van der Zaag, P. J., Franke, A., ... Brookes, A. J. (2011). Targeted enrichment of genomic DNA regions for next-generation sequencing. *Briefings in Functional Genomics*, *10*, 374–386. doi: 10.1093/bfgp/elr033.

Narzisi, G., O'Rawe, J. A., Iossifov, I., Fang, H., Lee, Y., Wang, Z., ... Schatz, M. C. (2014). Accurate de novo and transmitted indel detection in exome-capture data using microassembly. *Nature Methods*, *11*, 1033–1036. doi: 10.1038/nmeth.3069.

Porreca, G. J., Zhang, K., Li, J. B., Xie, B., Austin, D., Vassallo, S. L., ... Shendure, J. (2007). Multiplex amplification of large sets of human exons. *Nature Methods*, *4*, 931–936. doi: 10.1038/nmeth1110.

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, *29*, 24–26. doi: 10.1038/nbt.1754.

Shen, M.-J. R., Oliphant, A., Butler, S. L., Stuelpnagel, J. E., Chee, M. S., Kuhn, K. M., & Fan, J.-B. (2011). Multiplex nucleic acid reactions. U.S. Patent 7955794 B2.

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K. (2001). dbSNP: The NCBI database of genetic variation. *Nucleic Acids Research*, *29*, 308–311. doi: 10.1093/nar/29.1.308.

Snyder, M. W., Kircher, M., Hill, A. J., Daza, R. M., & Shendure, J. (2016). Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. *Cell*, *164*, 57–68. doi: 10.1016/j.cell.2015.11.050.

Srinivasan, M., Sedmak, D., & Jewell, S. (2002). Effect of fixatives and tissue processing on the content and integrity of nucleic acids. *The American Journal of Pathology*, *161*, 1961–1971. doi: 10.1016/S0002-9440(10)64472-0.

Tewhey, R., Warner, J. B., Nakano, M., Libby, B., Medkova, M., David, P. H., ... Frazer, K. A. (2009). Microdroplet-based PCR enrichment for large-scale targeted sequencing. *Nature Biotechnology*, *27*, 1025–1031. doi: 10.1038/nbt.1583.

Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, *14*, 178–192. doi: 10.1093/bib/bbs017.

Varley, K. E., & Mitra, R. D. (2008). Nested Patch PCR enables highly multiplexed mutation discovery in candidate genes. *Genome Research*, *18*, 1844–1850. doi: 10.1101/gr.078204.108.

Internet Resources

https://www.neb.com/nebnext-direct/nebnext-direct-for-target-enrichment
NEBNext Direct Web site.

https://www.neb.com/~/media/NebUs/Files/PDF%20FAQ/NEBNext%20Direct%20Cancer%20HotSpot%20Panel%20FAQ_4-27.pdf
NEBNext Direct FAQs: Includes additional information and resources including examples of command lines for data analysis.

http://support.illumina.com/downloads/miseq_system_user_guide_15027617.html
Illumina MiSeq System Guide.

http://support.illumina.com/downloads/miseq-reporter-user-guide-15042295.html
Illumina MiSeq Reporter Software Guide.

http://support.illumina.com/downloads/prepare_libraries_for_sequencing_miseq_15039740.html
MiSeq Denature and Dilute Libraries Guide.

http://broadinstitute.github.io/picard
Picard tools.

https://software.broadinstitute.org/gatk/gatkdocs/
Genome Analysis ToolKit (GATK).

http://bio-bwa.sourceforge.net/
Burrows-Wheeler Aligner (BWA) software package.

https://github.com/fulcrumgenomics/fgbio
Information on marking and working with duplicates.

https://www.ncbi.nlm.nih.gov/snp
The Single Nucleotide Polymorphism Database (dbSNP).

http://cancer.sanger.ac.uk
Cosmic database.