NEB® EXPRESSIONS

NEW ENGLAND BioLabs® Inc.

NEB offers several helpful interactive tools for your research and experimental design.

Double Digest Finder

NEBioCalculator™

Tm Calculator

NEBcutter®
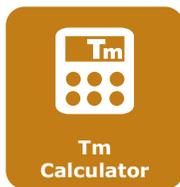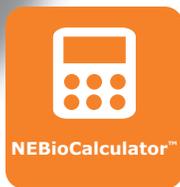
Explore the NEB toolbox at
www.neb.com/nebtools

Contents
Issue I, 2014

COVER PHOTO
Spring blossoms on the New England Biolabs campus. Photographed by Wendy Geary.
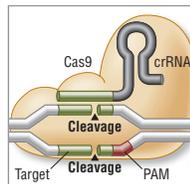
CRISPR/Cas9 and Targeted Genome Editing: A New Era in Molecular Biology

The development of efficient and reliable ways to make precise, targeted changes to the genome of living cells is a long-standing goal for biomedical researchers. Recently, a new tool based on a bacterial CRISPR-associated protein-9 nuclease (Cas9) from *Streptococcus pyogenes* has generated considerable excitement (1). This follows several attempts over the years to manipulate gene function, including homologous recombination (2) and RNA interference (RNAi) (3). RNAi, in particular, became a laboratory staple enabling inexpensive and high-throughput interrogation of gene function (4, 5), but it is hampered by providing only temporary inhibition of gene function and unpredictable off-target effects (6). Other recent approaches to targeted genome modification – zinc-finger nucleases [ZFNs, (7)] and transcription-activator like effector nucleases [TALENs (8)]– enable researchers to generate permanent mutations by introducing double-stranded breaks to activate repair pathways. These approaches are costly and time-consuming to engineer, limiting their widespread use, particularly for large scale, high-throughput studies.

Alex Reis, Ph.D., Bitesize Bio
Breton Hornblower, Ph.D., Brett Robb, Ph.D. and George Tzertzinis, Ph.D., New England Biolabs, Inc.

The Biology of Cas9

The functions of CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) and CRISPR-associated (Cas) genes are essential in adaptive immunity in select bacteria and archaea, enabling the organisms to respond to and eliminate invading genetic material. These repeats were initially discovered in the 1980s in *E. coli* (9), but their function wasn't confirmed until 2007 by Barrangou and colleagues, who demonstrated that *S. thermophilus* can acquire resistance against a bacteriophage by integrating a genome fragment of an infectious virus into its CRISPR locus (10).

Three types of CRISPR mechanisms have been identified, of which type II is the most studied. In this case, invading DNA from viruses or plasmids is cut into small fragments and incorporated into a CRISPR locus amidst a series of short repeats (around 20 bps). The loci are transcribed, and transcripts are then processed to generate small RNAs (crRNA – CRISPR RNA), which are

Figure 1. Cas9 *in vivo*: Bacterial Adaptive Immunity
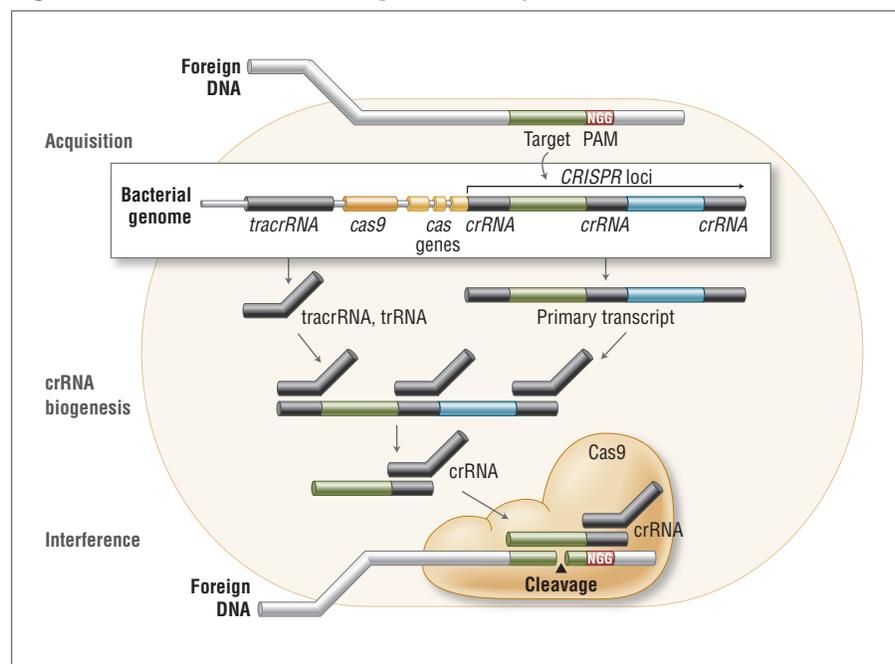
In the acquisition phase, foreign DNA is incorporated into the bacterial genome at the CRISPR loci. CRISPR loci is then transcribed and processed into crRNA during crRNA biogenesis. During interference, Cas9 endonuclease complexed with a crRNA and separate tracrRNA cleaves foreign DNA containing a 20-nucleotide crRNA complementary sequence adjacent to the PAM sequence. (Figure not drawn to scale.)

used to guide effector endonucleases that target invading DNA based on sequence complementarity (Figure 1) (11).

One Cas protein, Cas9 (also known as Csn1), has been shown, through knockdown and rescue experiments to be a key player in certain CRISPR mechanisms (specifically type II CRISPR systems). The type II CRISPR mechanism is unique compared to other CRISPR systems, as only one Cas protein (Cas9) is required for gene silencing (12). In type II systems, Cas9 participates in the processing of crRNAs (12), and is responsible for the destruction of the target DNA (11). Cas9's function in both of these steps relies on the presence of two nuclease domains, a RuvC-like nuclease domain located at the amino terminus and a HNH-like nuclease domain that resides in the mid-region of the protein (13).

Genome Editing Glossary

Cas = CRISPR-associated genes

Cas9, Csn1 = a CRISPR-associated protein containing two nuclease domains, that is programmed by small RNAs to cleave DNA

crRNA = CRISPR RNA

dCAS9 = nuclease-deficient Cas9

DSB = Double-Stranded Break

gRNA = guide RNA

HDR = Homology-Directed Repair

HNH = an endonuclease domain named for characteristic histidine and asparagine residues

Indel = insertion and/or deletion

NHEJ = Non-Homologous End Joining

PAM = Protospacer-Adjacent Motif

RuvC = an endonuclease domain named for an *E. coli* protein involved in DNA repair

sgRNA = single guide RNA

tracrRNA, trRNA = trans-activating crRNA

TALEN = Transcription-Activator Like Effector Nuclease

ZFN = Zinc-Finger Nuclease

To achieve site-specific DNA recognition and cleavage, Cas9 must be complexed with both a crRNA and a separate trans-activating crRNA (tracrRNA or trRNA), that is partially complementary to the crRNA (11). The tracrRNA is required for crRNA maturation from a primary transcript encoding multiple pre-crRNAs. This occurs in the presence of RNase III and Cas9 (12).

During the destruction of target DNA, the HNH and RuvC-like nuclease domains cut both DNA strands, generating double-stranded breaks (DSBs) at sites defined by a 20-nucleotide target sequence within an associated crRNA transcript (11, 14). The HNH domain cleaves the complementary strand, while the RuvC domain cleaves the non-complementary strand.

The double-stranded endonuclease activity of Cas9 also requires that a short conserved sequence, (2–5 nts) known as protospacer-associated motif (PAM), follows immediately 3´- of the crRNA complementary sequence (15). In fact, even fully complementary sequences are ignored by Cas9-RNA in the absence of a PAM sequence (16).

## Cas9 and CRISPR as a New Tool in Molecular Biology

The simplicity of the type II CRISPR nuclease, with only three required components (Cas9 along with the crRNA and trRNA) makes this system amenable to adaptation for genome editing. This potential was realized in 2012 by the Doudna and Charpentier labs (11). Based on the type II CRISPR system described previously, the authors developed a simplified two-component system by combining trRNA and crRNA into a single synthetic single guide RNA (sgRNA). sgRNA-programmed Cas9 was shown to be as effective as Cas9 programmed with separate trRNA and crRNA in guiding targeted gene alterations (Figure 2A).

To date, three different variants of the Cas9 nuclease have been adopted in genome-editing protocols. The first is wild-type Cas9, which can site-specifically cleave double-stranded DNA, resulting in the activation of the double-strand break (DSB) repair machinery. DSBs can be repaired by the cellular Non-Homologous End Joining (NHEJ) pathway (17), resulting in insertions and/or deletions (indels) which disrupt the targeted locus. Alternatively, if a donor template with homology to the targeted locus is supplied, the DSB may be repaired by the homology-directed repair (HDR) pathway allowing for precise replacement mutations to be made (Figure 2A) (17, 18).

Cong and colleagues (1) took the Cas9 system a step further towards increased precision by developing a mutant form, known as Cas9D10A, with only nickase activity. This means it cleaves only one DNA strand, and does not activate NHEJ. Instead, when provided with a homologous repair template, DNA repairs are conducted via the high-fidelity HDR pathway only, resulting in reduced indel mutations (1, 11, 19). Cas9D10A is even more appealing in terms of target specificity when loci are targeted by paired Cas9 complexes designed to generate adjacent DNA nicks (20) (see further details about "paired nickases" in Figure 2B and on page 5).

The third variant is a nuclease-deficient Cas9 (dCas9, Figure 2C) (21). Mutations H840A in the HNH domain and D10A in the RuvC domain inactivate cleavage activity, but do not prevent DNA binding (11, 22). Therefore, this variant can be used to sequence-specifically target any region of the genome without cleavage. Instead,

Figure 2. CRISPR/Cas9 System Applications
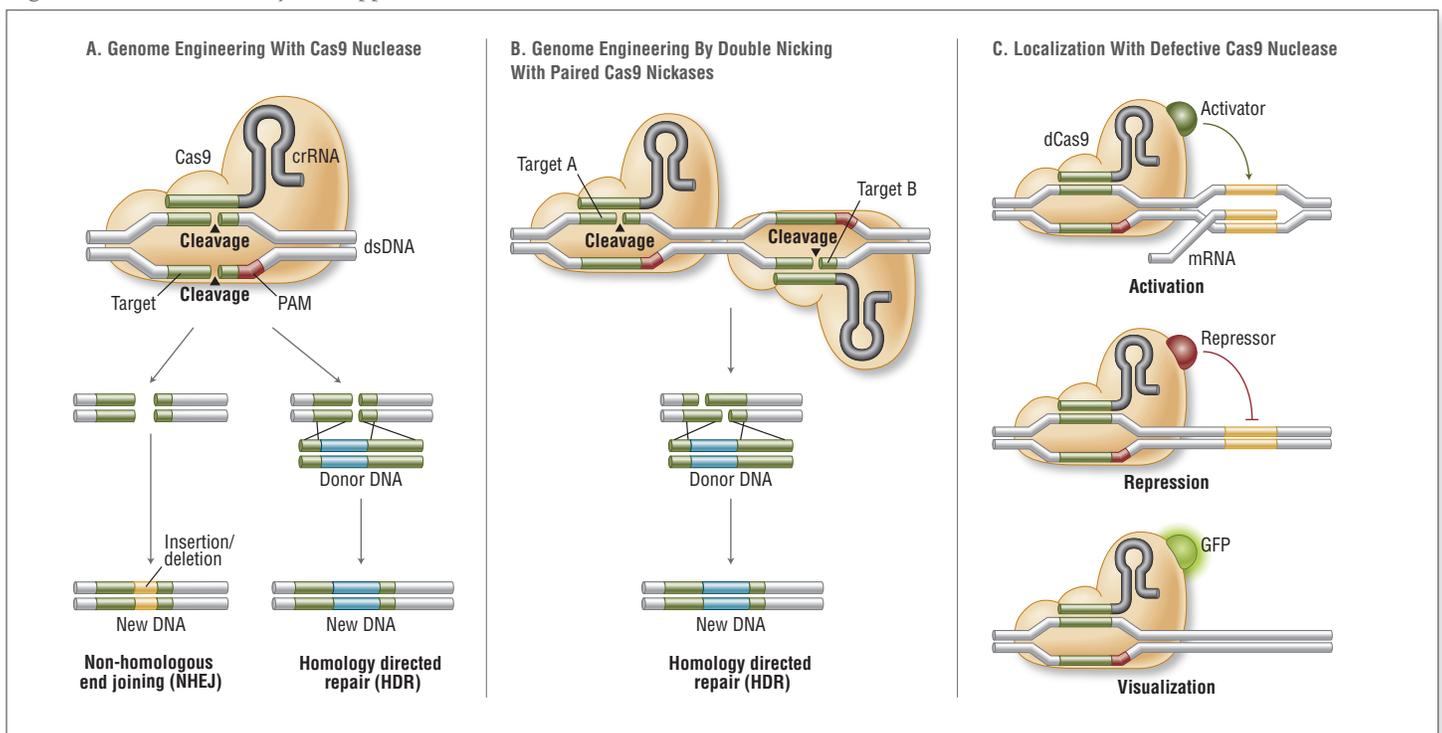
A. Wild-type Cas9 nuclease site specifically cleaves double-stranded DNA activating double-strand break repair machinery. In the absence of a homologous repair template non-homologous end joining can result in indels disrupting the target sequence. Alternatively, precise mutations and knock-ins can be made by providing a homologous repair template and exploiting the homology directed repair pathway.

B. Mutated Cas9 makes a site specific single-strand nick. Two sgRNA can be used to introduce a staggered double-stranded break which can then undergo homology directed repair.

C. Nuclease-deficient Cas9 can be fused with various effector domains allowing specific localization. For example, transcriptional activators, repressors, and fluorescent proteins.

by fusing with various effector domains, dCas9 can be used either as a gene silencing or activation tool (21, 23–26). Furthermore, it can be used as a visualization tool. For instance, Chen and colleagues used dCas9 fused to Enhanced Green Fluorescent Protein (EGFP) to visualize repetitive DNA sequences with a single sgRNA or non-repetitive loci using multiple sgRNAs (27).

Targeting Efficiency and Off-target Mutations

Targeting efficiency, or the percentage of desired mutation achieved, is one of the most important parameters by which to assess a genome-editing tool. The targeting efficiency of Cas9 compares favorably with more established methods, such as TALENs or ZFNs (8). For example, in human cells, custom-designed ZFNs and TALENs could only achieve efficiencies ranging from 1% to 50% (29–31). In contrast, the Cas9 system has been reported to have efficiencies up to >70% in zebrafish (32) and plants (33), and ranging from 2–5% in induced pluripotent stem cells (34). In addition, Zhou and colleagues were able to improve genome targeting up to 78% in one-cell mouse embryos, and achieved effective germline transmission through the use of dual sgRNAs to simultaneously target an individual gene (35).

A widely used method to identify mutations is the T7 Endonuclease I mutation detection assay (36, 37) (Figure 3). This assay detects heteroduplex DNA that results from the annealing of a DNA strand, including desired mutations, with a wild-type DNA strand (37).

Another important parameter is the incidence of off-target mutations. Such mutations are likely to appear in sites that have differences of only a few nucleotides compared to the original sequence, as long as they are adjacent to a PAM sequence. This occurs as Cas9 can tolerate up to 5 base mismatches within the protospacer region (36) or a single base difference in the PAM sequence (38). Off-target mutations are generally more difficult to detect, requiring whole-genome sequencing to rule them out completely.

Recent improvements to the CRISPR system for reducing off-target mutations have been made through the use of truncated gRNA (truncated within the crRNA-derived sequence) or by adding two extra guanine (G) nucleotides to the 5´ end (28, 37). Another way researchers have attempted to minimize off-target effects is with the use of "paired nickases" (20). This strategy uses D10A Cas9 and two sgRNAs complementary to the adjacent area on opposite strands of the target site (Figure 2B, page 4). While this induces DSBs in the target DNA, it is expected to create only single

Figure 3. T7 Endonuclease I Targeting Efficiency Assay
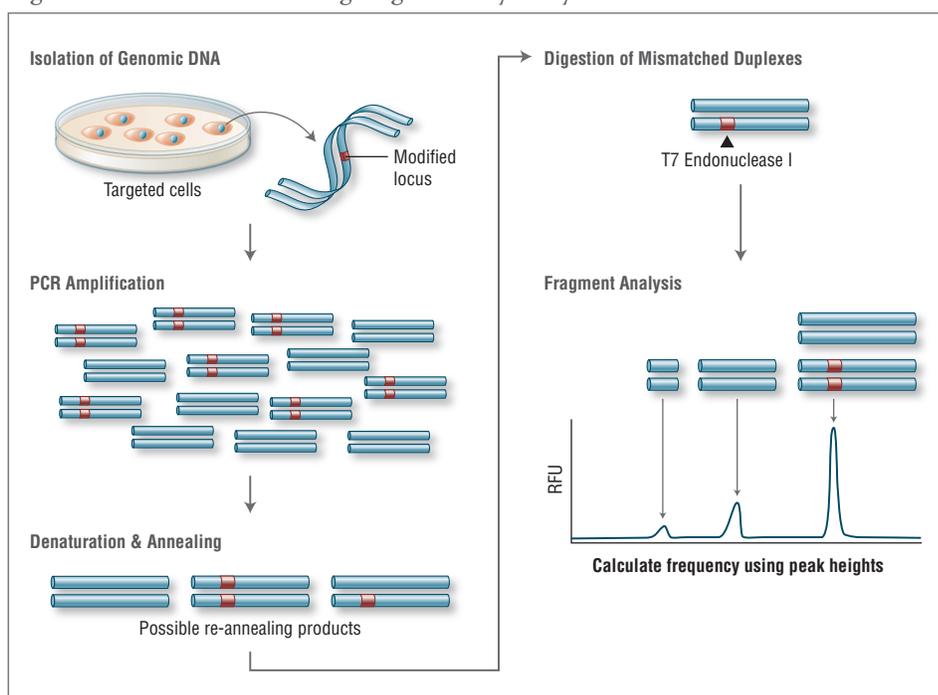
Genomic DNA is amplified with primers bracketing the modified locus. PCR products are then denatured and re-annealed yielding 3 possible structures. Duplexes containing a mismatch are digested by T7 Endonuclease I. The DNA is then electrophoretically separated and fragment analysis is used to calculate targeting efficiency.

nicks in off-target locations and, therefore, result in minimal off-target mutations.

By leveraging computation to reduce off-target mutations, several groups have developed web-based tools to facilitate the identification of potential CRISPR target sites and assess their potential for off-target cleavage. Examples include the CRISPR Design Tool (38) and the ZiFiT Targeter, Version 4.2 (39, 40).

Applications as a Genome-editing and Genome Targeting Tool

Following its initial demonstration in 2012 (9), the CRISPR/Cas9 system has been widely adopted. This has already been successfully used to target important genes in many cell lines and organisms, including human (34), bacteria (41), zebrafish (32), *C. elegans* (42), plants (34), *Xenopus tropicalis* (43), yeast (44), *Drosophila* (45), monkeys (46), rabbits (47), pigs (42), rats (48) and mice (49). Several groups have now taken advantage of this method to introduce single point mutations (deletions or insertions) in a particular target gene, via a single gRNA (14, 21, 29). Using a pair of gRNA-directed Cas9 nucleases instead, it is also possible to induce large deletions or genomic rearrangements, such as inversions or translocations (50). A recent exciting

development is the use of the dCas9 version of the CRISPR/Cas9 system to target protein domains for transcriptional regulation (26, 51, 52), epigenetic modification (25), and microscopic visualization of specific genome loci (27).

The CRISPR/Cas9 system requires only the re-design of the crRNA to change target specificity. This contrasts with other genome editing tools, including zinc finger and TALENs, where re-design of the protein-DNA interface is required. Furthermore, CRISPR/Cas9 enables rapid genome-wide interrogation of gene function by generating large gRNA libraries (51, 53) for genomic screening.

The future of CRISPR/Cas9

The rapid progress in developing Cas9 into a set of tools for cell and molecular biology research has been remarkable, likely due to the simplicity, high efficiency and versatility of the system. Of the designer nuclease systems currently available for precision genome engineering, the CRISPR/Cas system is by far the most user friendly. It is now also clear that Cas9's potential reaches beyond DNA cleavage, and its usefulness for genome locus-specific recruitment of proteins will likely only be limited by our imagination.

Alex Reis is the Channel Editor for Cloning Expression at Bitesizebio.com.

References

1. Cong L., et al. (2013) *Science*, 339, 819–823.
2. Capecchi, M.R. (2005) *Nat. Rev. Genet.* 6, 507–512.
3. Fire, A., et al. (1998) *Nature*, 391, 806–811.
4. Elbashir, S.Mm, et al. (2002) *Methods*, 26, 199–213.
5. Martinez, J., et al. (2003) *Nucleic Acids Res. Suppl.* 333.
6. Alic, N, et al. (2012) *PLoS One*, 7, e45367.
7. Miller, J., et al. (2005) *Mol. Ther.* 11, S35–S35.
8. Mussolino, C., et al. (2011) *Nucleic Acids Res.* 39, 9283–9293.
9. Ishino, Y., et al. (1987) *J. Bacteriol.* 169, 5429–5433.
10. Barrangou, R., et al. (2007). *Science*, 315, 1709–1712.
11. Jinek, M., et al. (2012) *Science,* 337, 816–821.
12. Deltcheva, E., et al. (2011) *Nature*, 471, 602–607.
13. Sapranauskas, R., et al. (2011) *Nucleic Acids Res.* 39, 9275–9282.
14. Nishimasu, H., et al. (2014) *Cell*, doi:10.1016/j.cell.2014.02.001
15. Swarts, D.C., et al. (2012) *PLoS One*, 7:e35888.
16. Sternberg, S.H., et al. (2014) *Nature*, doi:10.1038/nature13011.
17. Overballe-Petersen, S., et al. (2013) *Proc. Natl. Acad. Sci. U.S.A.* 110, 19860–19865.

18. Gong, C., et al. (2005) *Nat. Struct. Mol. Biol.* 12, 304–312.
19. Davis, L., Maizels, N. (2014) *Proc. Natl. Acad. Sci. U S A,* 111, E924–932.
20. Ran, F.A., et al. (2013) *Cell*, 154, 1380–1389.
21. Qi, L.S., et al. (2013) *Cell*, 152, 1173–1183.
22. Gasiunas, G., et al. (2012) *Proc. Natl. Acad. Sci. U S A,* 109, E2579–2586.
23. Maede, M.L., et al. (2013) *Nat. Methods*, 10, 977–979.
24. Gilbert, L.A., et al. (2013) *Cell*, 154, 442–451.
25. Hu, J., et al. (2014) *Nucleic Acids Res.* doi:10.1093/nar/gku109.
26. Perez-Pinera, P., et al. (2013) *Nat. Methods,* 10, 239–242.
27. Chen, B., et al. (2013) *Cell*, 155, 1479–1491.
28. Seung, W., et al. (2014) *Genome Res.* 24, 132–141.
29. Miller, J.C., et al. (2011). *Nat. Biotechnol.* 29, 143–148.
30. Mussolino, C., et al. (2011). *Nucleic Acids Res.* 39, 9283–9293.
31. Maeder, M.L., et al. (2008) *Mol. Cell,* 31, 294–301.
32. Hwang, W.Y., et al. (2013) *PLoS One*, 8:e68708.
33. Feng, Z., et al. (2013) *Cell Res.* 23, 1229–1232.
34. Mali, P., et al. (2013) *Science*, 339, 823–826.
35. Zhou, J., et al. (2014) *FEBS J.* doi:10.1111/febs.12735.
36. Fu, Y., et al. (2013) *Nat. Biotechnol.* 31, 822–826.

37. Fu, Y., et al. (2014) *Nat Biotechnol.* doi: 10.1038/nbt.2808.
38. Hsu, P.D., et al. (2013) *Nat. Biotechnol.* 31, 827–832.
39. Sander, J.D., et al. (2007) *Nucleic Acids Res.* 35, W599-605.
40. Sander, J.D., et al (2010) *Nucleic Acids Res.* 38, W462–468.
41. Fabre, L., et al. (2014) *PLoS Negl. Trop. Dis.* 8:e2671.
42. Hai, T., et al. (2014) *Cell Res.* doi: 10.1038/cr.2014.11.
43. Guo, X., et al. (2014) *Development,* 141, 707–714.
44. DiCarlo, J.E., et al. (2013) *Nucleic Acids Res.* 41, 4336–4343.
45. Gratz, S.J., et al. (2014) *Genetics,* doi:10.1534/genetics.113.160713.
46. Niu, Y., et al. (2014) *Cell*, 156, 836–843.
47. Yang, D., et al. (2014) *J. Mol. Cell Biol.* 6, 97-99.
48. Ma, Y., et al. (2014) *Cell Res.* 24, 122–125.
49. Mashiko, D., et al. (2014) *Dev. Growth Differ.* 56, 122–129.
50. Gratz, S.J., et al. (2013) *Fly*, 249.
51. Mali, P., et al. (2013) *Nat. Biotechnol.* 31, 833–838.
52. Cheng, A.W., et al. (2013) *Cell Res.* 23, 1163–1171.
53. Koike-Yusa, H., et al. (2013) *Nat. Biotechnol.* doi: 10.1038/nbt.2800.
54. Sander, J.D., and Joung, J.K. (2014) *Nat Biotechnol.* doi:10.1038/nbt.2842.

Tips for Planning Your CRISPR/Cas9 Experiments

The CRISPR/Cas9 genome editing technique is a powerful tool for researchers. However, several practical aspects should be carefully considered in order to achieve the best results from this system. Such considerations include: which promoter to use, whether to opt for wild-type of double nickase, how to achieve multiplexing and, perhaps most importantly, which vector to use.

For example, the chosen promoter may influence the range of target sites available (54). Using the U6 or T7 promoters requires a G or GG, respectively, at the 5´ end. Generating gRNAs with mismatches to the first two bases, or simply adding two guanines to the 5´ end, can reduce such restrictions.

Undoubtedly, the most important decision is to decide which vector to use. A variety of vectors have been validated for different cells and model organisms, and final application, from cutting or nicking to activating genes and screening libraries. Several groups have provided repositories of these plasmids, which are available through Addgene (www.addgene.org).

Examples of NEB products that can be used to support CRISPR workflows are shown below. Additional products to support template construction are shown on page 7.

Featured NEB Products Supporting CRISPR Workflows

PRODUCT NAME	CRISPR/Cas9 APPLICATION	NEB #	SIZE
NEW Cas9 Nuclease, *S. pyogenes*	Central component in the generation of CRISPR-based immunity – catalyzes site-specific cleavage of double-stranded DNA	M0386S	50 rxns
Q5® Site-directed Mutagenesis Kit (with or without competent cells)	Insertion of target sequence into the Cas9-sgRNA construct	E0554S/E0552S	10 rxns
Q5 High-fidelity DNA Polymerases	High-fidelity construct generation for use with CRISPR workflows	Multiple	Multiple
Gibson Assembly® Master Mix	Single-tube, isothermal generation of the Cas9-sgRNA construct	E2611S/L	10/50 rxns
Gibson Assembly® Cloning Kit	Single-tube, isothermal generation of the Cas9-sgRNA construct	E5510S	10 rxns
T7 Endonuclease I	Determination of the targeting efficiency of genome editing protocols	M0302S/L	250/1,250 units
HiScribe™ T7 High Yield RNA Synthesis Kit	Generation of sgRNA	E2040S	50 rxns
HiScribe T7 Quick High Yield RNA Synthesis Kit	Generation of sgRNA	E2050S	50 rxns

ONLINE RESOURCES

Plasmid Repositories:
http://www.addgene.org

CRISPR-gRNA Design Tools:
http://crispr.mit.edu
http://zifit.partners.org/ZiFiT/
http://www.e-crisp.org/E-CRISP/designcrispr.html
https://chopchop.rc.fas.harvard.edu/

Online Forums:
https://groups.google.com/forum/#!forum/crispr

Organism-specific Resources:
http://wormcas9hr.weebly.com
http://www.flyrnai.org

FEATURED APPLICATION

sgRNA Template Construction for CRISPR/Cas9 Genome Editing

Plasmids containing sgRNA sequences can be constructed using a variety of methods. Common to each is the requirement to introduce an approximately 20 base target sequence downstream of a promoter. Guide RNA templates can then be used as templates for *in vitro* transcription or directly introduced. The figure below shows common strategies and the accompanying NEB products.
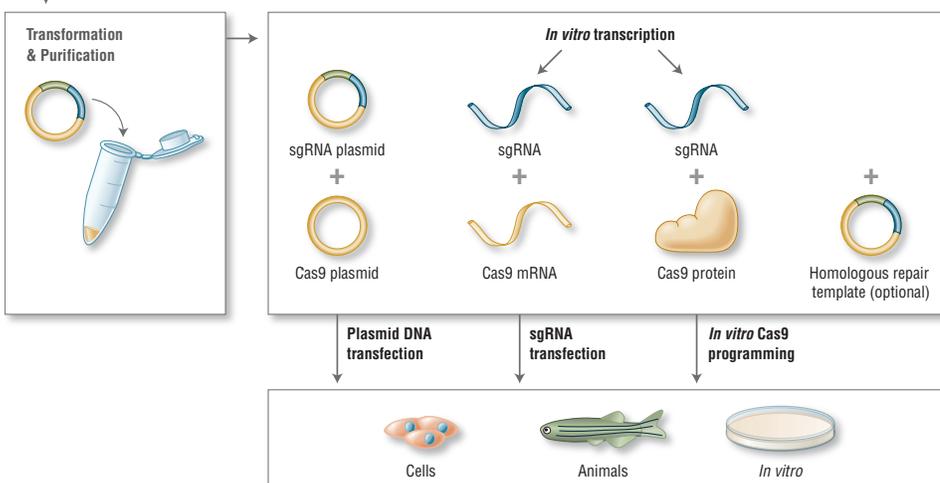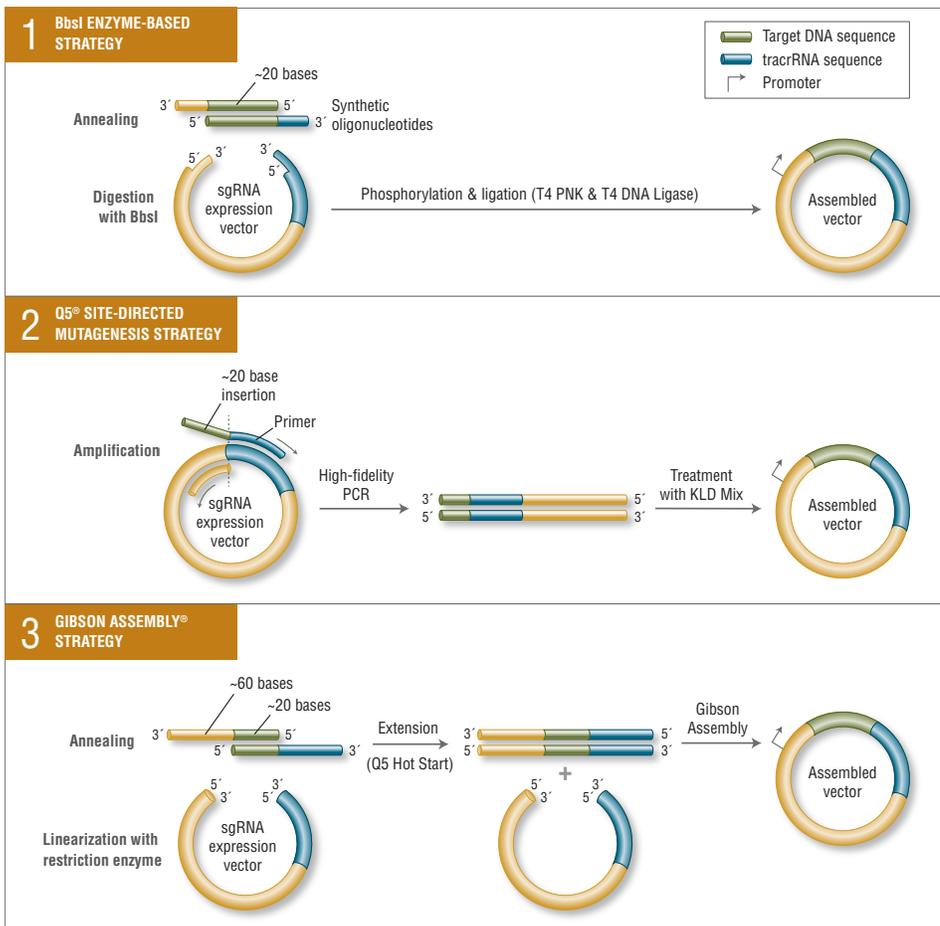
Ordering Information

PRODUCT	NEB #
BbsI	R0539S/L
T4 DNA Ligase	M0202S/M/L/T
T4 Polynucleotide Kinase	M0201S/L

PRODUCT	NEB #
Q5 Site-Directed Mutagenesis Kit	E0554S
Q5 Site-Directed Mutagenesis Kit (Without Competent Cells)	E0552S

PRODUCT	NEB #
Q5 Hot Start High-Fidelity 2X Master Mix	M0494S/L
Gibson Assembly Cloning Kit	E5510S
Gibson Assembly Master Mix	E2611S/L

PRODUCT	NEB #
NEB 5-alpha Competent *E. coli* (High Efficiency)	C2987P/I/H
NEB 10-beta Competent *E. coli* (High Efficiency)	C3019I/H
HiScribe T7 High Yield RNA Synthesis Kit	E2040S
HiScribe T7 Quick High Yield RNA Synthesis Kit	E2050S
Anti-Reverse Cap Analog 3′-*O*-Me-m7G(5′) ppp(5′)G	S1411S/L
Vaccinia Capping System	M2080S

SplintR™ Ligase

SplintR Ligase, also known as PBCV-1 DNA Ligase or *Chlorella* virus DNA Ligase, efficiently catalyzes the ligation of adjacent, single-stranded DNA oligonucleotides, splinted by a complementary RNA. The robust activity of the enzyme and its affinity for RNA-splinted DNA substrates (apparent Km ca. 1 nM) enable sub-nanomolar detection of unique RNA species within a complex mixture, making SplintR Ligase a superior choice for demanding RNA detection technologies.

Figure 1. RNA-splinted DNA Ligation with SplintR Ligase
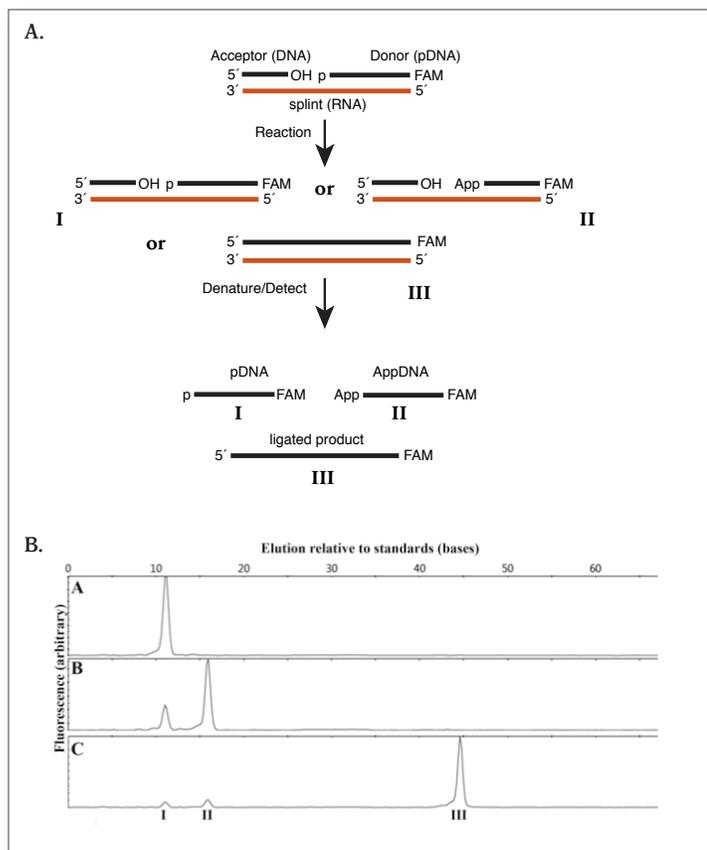
Ligation of DNA splinted by RNA: (A) Outline of the ligation assay: a 5´-phosphorylated, 3´-FAM labeled DNA "donor" oligonucleotide and an unmodified DNA "acceptor" oligonucleotide are annealed to a complementary RNA splint. This substrate is reacted with a ligase to form a mixture of unreacted starting material (I), adenylylated DNA (II), and ligated product (III). These products are denatured, separated by capillary electrophoresis and detected by fluorescence. (B) Ligation of the RNA-splinted substrate in SplintR Ligase Reaction Buffer for 15 minutes at 25°C with (a) no enzyme, (b) 1 μM T4 DNA Ligase and (c) 100 nM SplintR Ligase. Indicated peaks correspond to starting pDNA (I), AppDNA (II) and ligated product (III), as determined by co-elution with synthetically prepared standards.

Ordering Information

PRODUCT	NEB #	SIZE
SplintR Ligase	M0375S/L	1,250/6,250 units

FAQ Spotlight

Q: **Can T4 DNA Ligase be used to ligate DNA on an RNA Splint?**

A: *T4 DNA Ligase has been reported to perform this ligation under conditions of very low ATP concentration (10 μM). SplintR Ligase has been shown to perform this ligation with shorter incubation times, significantly higher ligation yields (approaching 100%) and is not inhibited by ATP.*

TECHNICAL TIPS

Getting Started with RNA-Seq

While microarrays have historically been used for RNA profiling, this technique suffers from significant shortcomings, including low accuracy and low sensitivity. In contrast, RNA sequencing, commonly referred to as RNA-seq, can deliver unbiased information about the transcriptome. This method allows the identification of exons and introns, as well as identification of the 5´ and 3´ ends of genes.

Step 1. Creating a cDNA Library

Several methods can be used to generate an RNA-seq library, and the details of these methods will be dependent on the platform used for high-throughput sequencing. However, there are common steps:

• **Removal:** Since the majority of RNA molecules present in a cell are ribosomal RNA (rRNA), which are generally not of interest, they should be removed before making a library of the RNA of interest. Two popular options for this step are:

– **mRNA isolation:** This involves targeting the polyadenylated (poly(A)) tails to ensure that non-coding RNA is separated from polyadenylated transcripts.

– **rRNA depletion:** Total RNA may be depleted of rRNA using a number of methods, most of which involve hybridization of oligos to the rRNA followed by removal. rRNA depletion enables subsequent sequencing of all non-rRNA molecules, and is not limited to intact mRNA molecules.

• **Fragmentation:** Shearing generates fragments of an appropriate size for sequencing, and is accomplished by an RNA fragmentation step prior to reverse transcription, rather than by fragmentation of cDNA.

• **Reverse transcription and second-strand cDNA synthesis:** RNA is converted into a single-stranded cDNA library via random or oligo(dT) primers and a reverse transcriptase. The resulting cDNA is then converted into double-stranded cDNA by a DNA polymerase.

• **End repair, dA-Tailing and Adaptor Ligation:** End repair of the cDNA library followed by optional dA-tailing (depending on the sequencing platform to be used) is followed by ligation to 3´ & 5´ adaptors. The library is then ready for amplification and sequencing.

Step 2. Sequencing and analysis

High-throughput sequencing technologies generate a large number of sequence reads from a library of DNA fragments. Sequence reads are mapped against a reference genome. Software packages are available for short-read alignment, and specialized algorithms for transcriptome alignment have been developed, including TopHat and Cufflinks.

Visit **NEBNext.com** to find the full list of products available for this application.

LEARN MORE WITH OUR VIDEO TECH TIPS AT NEBNEXT.COM

Tips for Optimizing RNA Inputs

12 Quick Tips for NGS Library Prep

IMPORTANT FACTORS TO CONSIDER WHEN PERFORMING RNA-SEQ

Several methods are currently available for library preparation for RNA-seq, many of which offer simplified protocols and improved yields. However, the quality and accurate quantitation of input RNA still remains critical to ensuring successful cDNA synthesis and libraries. The following are some important factors to consider:

• **Quality of RNA sample:** High-quality RNA is essential for successful cDNA library preparation, and care should be taken when handling RNA samples. As RNA is prone to degradation by ribonucleases, an RNase-free environment is essential. For tips on avoiding RNase contamination, visit **www.neb.com/ RibonucleaseContamination**. The RNA Integrity Number (RIN), as determined by the Agilent Bioanalyzer®, is a useful measurement of RNA quality, and a RIN of 7 or higher is ideal.

• **Quantity of RNA needed:** Protocols suitable for input amounts in the low ng range are now available, and accurate quantitation of input RNA is important. Contaminants present in the sample may affect quantitation; for example, free nucleotides or other organic compounds routinely used to extract RNA will also absorb UV light near 260 nm, and will result in an over-estimation of RNA concentration when spectrophotometric methods are used.

• **Strand-specific or non-directional libraries:** Unlike standard methods, directional (strand-specific) protocols for sequencing RNA provide information on the DNA strand from which the RNA strand was transcribed. This is useful for many reasons including:

– Identification of antisense transcripts

– Determination of the transcribed strand of noncoding RNAs

– Determination of expression levels of coding or noncoding overlapping transcripts.

• **RNA expression uniformity:** It is important to be aware that results are tissue-specific and time-dependent. Gene expression is not uniform throughout an organism's cells, and it is strongly dependent on the tissue being assessed. In addition, gene expression levels change over the lifetime of a cell.

Automated size selection of NEBNext Small RNA libraries with the Sage Pippin Prep

Daniela Munafo, Ph.D.[1], Sadaf Hoda, Ph.D.[2], Laurence Ettwiller, Ph.D.[1], Brad Langhorst, Ph.D.[1], Eileen Dimalanta, Ph.D.[1], Fiona Stewart, Ph.D.[1], Chris Boles, Ph.D.[2]

Introduction

One of the fastest growing areas of biological research is regulatory small RNA structure, processing, and function. Next generation sequencing (NGS) is currently the method of choice for studying the variety and expression of small RNAs.

A common problem in NGS methods for small RNAs is contamination from adaptor-dimer artifacts, because these artifacts are very close in size to the small RNA library elements. To address this problem, the NEBNext Small RNA Library Prep Kits from New England Biolabs use specially engineered RNA ligases and optimized workflows, as well as a novel technology to dramatically reduce the formation of adaptor-dimer artifacts during library construction. Since the workflow uses total RNA as the starting material, it is beneficial to perform a final size selection step on the amplified libraries. NEB has previously validated size selection methods using AMPure XP® beads and manual preparative gel electrophoresis on a 6% polyacrylamide gel. Here we validate the use of Pippin Prep 3% gel cassettes for size selection of NEBNext Small RNA libraries.

Results

Library profile before and after size selection

The experiments were designed to compare NEBNext Small RNA library quality using two gel-based protocols: manual preparative electrophoresis on 6% PAGE gels, and automated agarose gel electrophoresis on the Pippin Prep.

Six human brain small RNA libraries were prepared from total RNA and indexed using different barcodes. Three libraries were PAGE-size-selected (Figure 1B) and the other three libraries were size-selected on the Pippin Prep using a 3% dye-free agarose cassette (Figure 1C). Both size selected libraries have only a single peak at ~149 bp, with minimal contamination from smaller or larger species. Library yields from the size-selected libraries using the manual gel and Pippin Prep were similar (~20% of the input material is recovered; data not shown).

MiSeq® Sequencing and Data Analysis

The 6 barcoded libraries were pooled in equimolar concentrations, loaded onto a MiSeq Reagent v2 Kit at 8 pM final concentration, and sequenced on a MiSeq instrument (SE; 1X 36bp; 2.5 million of reads/library).

Adapter trimming and filtering reads by length

Reads were adaptor-trimmed and filtered by length. Reads shorter than 15 nucleotides were discarded. A high percentage of reads (>90% total reads for both size selection methods) passed the length filtering. The NEBNext Small RNA library prep optimized workflow prevents adaptor-dimer formation, therefore only a minimal percentage of reads did not contain insert.

Mapping Rate

A high percentage of trimmed reads (>66%) mapped to human miRNAs present in miRBase for both gel and Pippin Prep size selection methods.

The vast majority of trimmed reads (98%) mapped either to the human genome or to miRBase. From the reads that mapped to human, an average of 28% overlapped at least one exonic feature (miRNA, lincRNA, pseudogene, snRNA, protein coding RNA, snoRNA, rRNA, sense_intronic, mt-tRNA, or antisense RNAs). Some of the reads that did not map to miRBase overlapped with a putative microRNA, increasing the number of reads mapping to various microRNAs (from miRBase or gencode annotations) to more than 68% of total trimmed reads (data not shown).

miRNA Expression Analysis

miRNA expression correlation of the 500 most abundant miRNAs between PAGE and Pippin prep size-selected libraries was excellent [(R^2 = 0.99805) (y = 1.0558x + 15.089)]. This data indicates that small RNA expression levels were not biased due to different size selection methods (data not shown).

Conclusions

The experiments reported here show that NEBNext Small RNA libraries produced using the Sage Pippin Prep equal or exceed the quality of libraries size selected using the manual gel procedure. The benefits are seen in library purity.

Two other key features of the Pippin Prep that are not shown in the present experiments are reproducibility and ease-of-use. The manual gel procedure has many individual manual steps that are time-consuming and extremely difficult to perform reproducibly. In contrast, the Pippin Prep procedure requires no manual manipulations, except for gel loading. The entire separation and size-selection process is controlled by the onboard computer. This removes all opportunity for the user to introduce variability into the process.

Hands-on effort and time for the Pippin Prep procedure is 15-20 minutes to set up and load a cassette, and less than 5 minutes to remove samples at the end of the run. Run time for the NEBNext Small RNA libraries in the 3% internal standard cassette is a little over one hour. Up to five samples can be run per cassette.

In summary, we have developed an optimized Pippin Prep protocol for use with the NEBNext Small RNA Library Prep Kits. The Pippin Prep protocol provides all of the sequence quality benefits of the standard manual gel protocol, but with greatly enhanced ease-of-use and reproducibility.

Visit **NEBNext.com** to download the full application note.

Figure 1. Library size distribution before and after size selection.
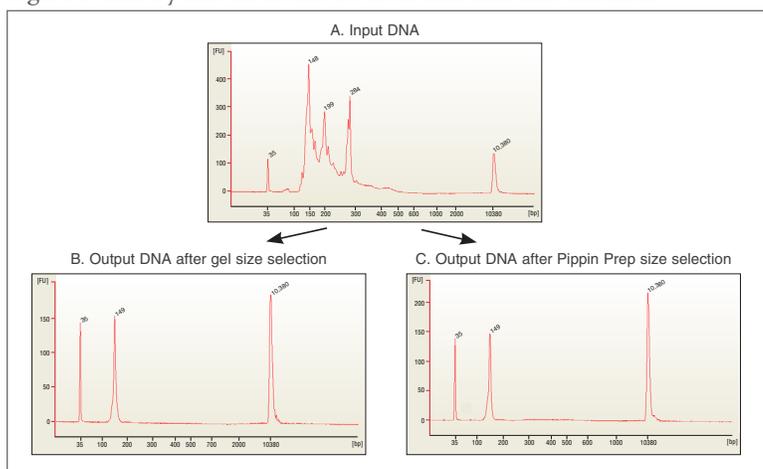
Bioanalyzer traces from NEBNext Human Brain miRNA libraries before size selection (A) and after size selection on a 6% polyacrylamide gel (B) or a 3% agarose (dye-free) cassette for Pippin Prep (C). Instrument Program Mode = Range; Start (bp) = 110 and End (bp) = 160 (C)

[1] New England Biolabs, Ipswich, MA; [2] Sage Science, Beverly, MA.

FEATURED PRODUCTS

NEBNext Reagents for Small RNA Library Preparation

Sequencing of Small RNA requires a different workflow than sequencing of other types of nucleic acids. NEBNext products for Small RNA library prep provide oligos, enzymes, buffers and protocols for the preparation of libraries from multiple types of Small RNAs. Our novel Small RNA workflow has been optimized to minimize adaptor-dimers while producing high-yield, high-diversity libraries.

Adaptors and primers are included in the Small RNA kits, and multiplexing options are available. The Multiplex Kits contain index primers, and the Multiplex-Compatible kit enables use with your own barcode system.

Small RNA Library Prep Workflow

3´ Adaptor Ligation → Primer Hybridization → 5´ Adaptor Ligation → First Strand Synthesis → PCR Enrichment → Size Selection

Input Amount 100 ng – 1 µg
Hands-On Time 30 min
Total Time 6 hrs

Ordering Information

PRODUCT	NEB #	SIZE
NEBNext Multiplex Small RNA Library Prep Set for Illumina (Set 1)	E7300S/L	24/96 rxns
NEBNext Multiplex Small RNA Library Prep Set for Illumina (Set 2)	E7580S/L	24/96 rxns
NEBNext Small RNA Library Prep Set for Illumina (Multiplex Compatible)	E7330S/L	24/96 rxns

ADVANTAGES

- Novel protocol for Small RNA
- Higher yield
- Substantially reduced adaptor-dimer formation
- Suitable for methylated and unmethylated small RNAs
- Broad range of input amounts (100 ng-1 µg)
- Minimal hands-on time
- Convenient format – Reagents, adaptors and primers are included. Multiplex option is available.
- Functional validation – Each product is functionally validated by library preparation from a standard reference RNA followed by Illumina® sequencing.
- Stringent quality controls – Additional QCs ensure maximum quality and purity.
- Suitable for library preparation for all Illumina sequencing instruments
- Low price per reaction

Sustainability at NEB

Did you know:

• NEB employs sustainable practices across its campus, including the LEED-certified laboratory building. These initiatives include:
  ◦ an extensive café composting program, which diverts as much as 75% of café waste away from landfills
  ◦ a comprehensive recycling program for paper goods, plastics, batteries and gel cooler packs
  ◦ carpooling/ride-share, for employees who live outside of the local area

• Thanks to NEB's on-site wastewater treatment facility, 100% of the campus' wastewater is treated and purified via a Solar Aquatics Systems®, which leaves the water suitable for reuse or watershed recharge – minimizing NEB's imprint on the local environment.

• To ensure its products stay at the optimal temperature during shipment, NEB make use of Koolit® Gel Packs from Cold Chain Technologies. These can be disposed of in an environmentally responsible manner. To learn how, please refer to Cold Chain Technologies' website www.coldchaintech.com/products/item/koolit-gel-packs.

• In 2013, Harvard University piloted a "Reuse Room" where researchers can deposit specific items for recycling and reuse, including NEB styrofoam coolers and ice packs. To learn more, visit www.laboratoryequipment.com/articles/2014/04/collaborative-sustainability.

"Stephanie (NEB) has been a wonderful ally and supporter and we appreciate her willingness to work with us on this exciting pilot program. We also appreciate the shared values of environmental conservation between our organizations – it's refreshing to see businesses that sell products on our campus take a sincere and personal interest in this topic."
– Jamie, Faculty of Arts and Sciences (FAS) Green Program Coordinator, Harvard University

To learn more about other sustainability practices at NEB, visit **www.neb.com/sustainability**

NEW ENGLAND BioLabs Inc.

New England Biolabs, Inc., 240 County Road, Ipswich, MA 01938-2723

Campus-Class™ Mail
Intra-Mail Network

FSC
www.fsc.org
MIX
Paper from responsible sources
FSC® C007074

DNA CLONING
DNA AMPLIFICATION & PCR
EPIGENETICS
RNA ANALYSIS
LIBRARY PREP FOR NEXT GEN SEQUENCING
PROTEIN EXPRESSION & ANALYSIS
CELLULAR ANALYSIS

www.neb.com

USA
New England Biolabs, Inc.
Telephone: (978) 927-5054
Toll Free: (U.S. Orders) 1-800-632-5227
Toll Free: (U.S. Tech) 1-800-632-7799
Fax: (978) 921-1350
info@neb.com

Canada
New England Biolabs, Ltd.
Toll Free: 1-800-387-1095
info.ca@neb.com

China, People's Republic
New England Biolabs (Beijing), Ltd.
Telephone: 010-82378265/82378266
info@neb-china.com

France
New England Biolabs France
Telephone : 0800 100 632
info.fr@neb.com

Germany & Austria
New England Biolabs GmbH
Free Call 0800/246 5227 (Germany)
Free Call 00800/246 52277 (Austria)
info.de@neb.com

Japan
New England Biolabs Japan, Inc.
Telephone: +81 (0)3 5669 6191
info@neb-japan.com

Singapore
New England Biolabs Pte. Ltd.
Telephone +65 6776 0903
sales.sg@neb.com

United Kingdom
New England Biolabs (UK), Ltd.
Call Free 0800 318486
info.uk@neb.com

Try NEBioCalculator™, our latest online tool, for scientific calculations and conversions.