

Illumina TruSeq DNA Adapters De-Mystified

James Schiemer

The key to sequencing random fragments of DNA is by the addition of short nucleotide sequences which allow any DNA fragment to:

- 1) Bind to a flow cell for next generation sequencing
- 2) Allow for PCR enrichment of adapter-ligated DNA fragments only
- 3) Allow for indexing or “barcoding” of samples so multiple DNA libraries can be mixed together into 1 sequencing lane (known as *multiplexing*)

Though you can easily buy kits and add these adapters on without any knowledge of how they work or what their structure is, it is enormously beneficial to know the theory behind it so that you can avoid tragic and expensive mistakes, as well as design your own adapters and primers if necessary.

First, let’s have a look at the Sequence of the adapters individually:

TruSeq Universal Adapter:

5 AATGATACGGCGACCAACGAGATCTACACTCTTCCCTACACGACGCTCTTCCGATCT 3

TruSeq Indexed Adapter

5 GATCGGAAGAGCACACGTCTGAACTCCAGTCAC-NNNNNN-ATCTCGTATGCCGTCTTCTGCTTG 3

Here “N” is any nucleotide, and the 6 of them together are a unique sequence which can readily be identified as unique to 1 library.

It is important to know how these sequences are oriented around your DNA fragments, how they bind to the flow cell, and how the indexes are read.

1. Preparation of Adapters

First, the preparation of the adapters is important. If you are ordering your own you will need to know the proper way to make them, and exactly how they align.

The TruSeq Universal adapter and the Indexed adapter have a very short region where they are complimentary.

Examining the 5'→3' of the TruSeq Universal Adapter and the 3'-5' Sequence of the Indexed adapter we see:

```

5      AATGATACGGCGACCAACGAGATCTACACTCTTTCCCTACACGAC--GCTCTTCCGATC*T 3
3  GTTCGTCTTCTGCCGTATGCTCTA(INDEX)CACTGACCTCAAGTCTGCACA--CGAGAAGGCTAG*P 5

```

The only complementary sequence is the last 12 nucleotides. In order to take these individual oligonucleotides and anneal them you simply add equimolar stocks of each (for example, 100uM of each oligo) into an eppendorf tube, heat the mixture on a thermomixer or heating block at 95C for 15-30 minutes, lowering the temperature to 70C for 15 minutes, and then allowing to cool to RT. This should have given the oligos more than enough time to denature and anneal to one another and form these adapters with the “floppy” overhangs.

You will also notice that when ordering these oligos you will absolutely need a 5' phosphate group on your indexed adapter, as well as a recommendation that you have a phosphorothioate bond between the 3' end C and T nucleotides on the TruSeq Universal Adapter. The 5' phosphate group is necessary for ligation of the adapter to your DNA fragment of interest, whereas the phosphorothioate bond allows a very strong bond between the C and T overhang. This is critical, as any exonuclease activity could remove the T nucleotide, which is necessary for annealing to the 3' A overhang created during library preparation.

2. Orientation of adapters around the DNA fragment of interest prior to enrichment

Just for an idea of how the DNA fragment of interest is flanked by the adapters I've created an image below.

```

LEFT OF INSERT
5      AATGATACGGCGACCAACGAGATCTACACTCTTTCCCTACACGAC--GCTCTTCCGATC*T 3
3  GTTCGTCTTCTGCCGTATGCTCTA(INDEX)CACTGACCTCAAGTCTGCACA--CGAGAAGGCTAG*P 5

RIGHT OF INSERT
5  P*GATCGGAAGAGC--ACACGTCTGAACTCCAGTCAC(INDEX)ATCTCGTATGCCGTCTTCTGCTTG 3
3  T*CTAGCCTTCTCG--CAGCACATCCCTTTCTCACATCTAGAGCCACCAGCGGCATAGTAA 5

```

Just to reiterate, there is a 12 base sequence in which the adapters are complementary, followed by a longer string of nucleotides which are hanging off.

Illumina TruSeq Adapters Demystified Rev. A, © 2011 Tufts University Core Facility
 Oligonucleotide sequences © 2007-2011 Illumina, Inc. All rights reserved. Derivative works
 created by Illumina customers are authorized for use with Illumina instruments and products
 only. All other uses are strictly prohibited.

For more information please go to genomics.med.tufts.edu or email tucf-htseq@tufts.edu

3. Orientation of adapters around the DNA fragment of interest during/after Enrichment

The PCR primers used in our lab have the sequence:

PCR Primer 1.0

5 AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGA 3

PCR Primer 2.0

5 CAAGCAGAAGACGGCATACGAGAT 3

PCR primer 1.0 can be directly read as the first 44 bases of the TruSeq Universal adapter, and therefore binds to the complement of that sequence during the PCR process.

5 AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT 3

PCR primer 2.0 can be read as the reverse complement of the last 24 bases of the TruSeq Indexed adapter.

5 GATCGGAAGAGCACACGTCTGAACTCCAGTCAC-NNNNNN-ATCTCGTATGCCGTCTTCTGCTTG 3

With PCR amplification of both the Sense or antisense strands you will no longer have a “floppy” overhang in the product, but instead you will have double stranded DNA all the way throughout looking something like this:



- Universal Adapter
- DNA Fragment of Interest
- Indexed Adapter
- 6 Base Index Region

Illumina TruSeq Adapters Demystified Rev. A, © 2011 Tufts University Core Facility
Oligonucleotide sequences © 2007-2011 Illumina, Inc. All rights reserved. Derivative works created by Illumina customers are authorized for use with Illumina instruments and products only. All other uses are strictly prohibited.

For more information please go to genomics.med.tufts.edu or email tucf-htseq@tufts.edu

4. Binding of Enriched product to a flow cell

When preparing to sequence the DNA, Illumina's protocol calls for denaturing of the DNA with 2N NaOH. This allows for single stranded DNA to bind onto the flow cell, and undergo bridge amplification (not going to be discussed here).

One very crucial part of the adapter design is that it binds to the flow cell in ssDNA form, and it does so using the following regions:

TruSeq Universal Adapter

5 AATGATACGCGACCAACGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT 3

Portion Bound To Flow Cell:

5 AATGATACGCGACCAACGA

TruSeq Indexed Adapter

5 GATCGGAAGAGCACACGTCTGAACTCCAGTCAC--NNNNNN--ATCTCGTATGCCGTCTTCTGCTTG 3

CGTATGCCGTCTTCTGCTTG 3

Illumina TruSeq Adapters Demystified Rev. A, © 2011 Tufts University Core Facility
 Oligonucleotide sequences © 2007-2011 Illumina, Inc. All rights reserved. Derivative works
 created by Illumina customers are authorized for use with Illumina instruments and products
 only. All other uses are strictly prohibited.

For more information please go to genomics.med.tufts.edu or email tucf-htseq@tufts.edu

5. Sequencing

The actual sequencing of the DNA is pretty straight forward, assuming you managed to get through the material prior to this. The Read 1 primer is essentially the sequence of the TruSeq Universal Adapter, excluding of course the portion which binds to the flow cell.

TruSeq Universal Adapter

5 AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT 3

Read 1

5 ACACTCTTTCCCTACACGACGCTCTTCCGATCT 3

The read two primer, for paired end reads, is essentially the same idea, only reading in the opposite direction.

The index read primer for any multiplexed samples simply binds to all the sequence prior to the 6BP barcode, and then reads these off one by one so that each cluster can be identified as a unique sample within the flow cell lane.

5 GATCGGAAGAGCACACGTCTGAACTCCAGTCAC--NNNNNN--ATCTCGTATGCCGTCTTCTGCTTG 3
 GATCGGAAGAGCACACGTCTGAACTCCAGTCAC

General Warning

If you plan to create your own adapter sequences, indexes, and primers it is essential to KNOW WHAT YOU ARE DOING. Without everything exactly right you are liable to get no data from your sequencing run, and have spent a lot of time and money on nothing. It is critical that you are sure you know what you're doing, or use a protocol that has been tested and proven to work. Illumina will not officially support any method other than its own, and we strongly advise using their method, or at the very least, make only minor adjustments.