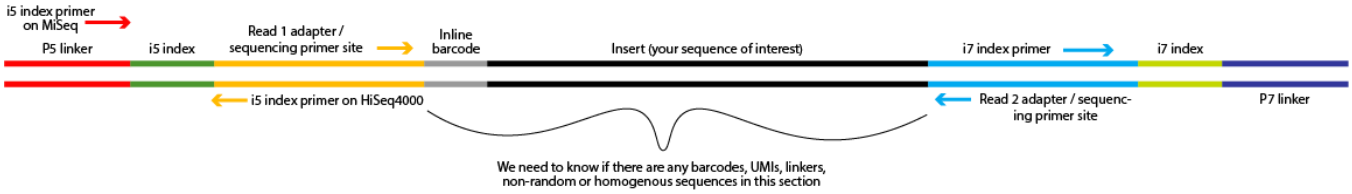
# ILLUMINA LIBRARY STRUCTURE




**All sequencing** of any type on the MiSeq or HiSeq4000 MUST HAVE FULL-LENGTH P5 and P7 sequences. (Some of the old small RNA libraries and alternative genomic library constructions used a partial P7; this is not supported by the HiSeq2000 PE, HiSeq4000, or MiSeq, and they should not be used to build new libraries. The P5 and P7 linkers are how the library is hybridized to the attachment sites on Illumina flowcells.

```
P5: 5' AAT GAT ACG GCG ACC ACC GA 3'

P7: 5' CAA GCA GAA GAC GGC ATA CGA GAT 3'
```

When checking your libraries by cloning and standard Sanger sequencing, this is the arrangement:

| | |
|---|---|
| P5--Read1primer--INSERT--Index/Read2(rc)primer—i7 index--P7(rc) | TruSeq (single index) library structure |
| P5--i5 index--Read1primer--INSERT--Index/Read2(rc)primer--i7 index--P7(rc) | Nextera or TruSeq HT (dual index) library |
| P5--Read1primer--INSERT--Read2primer(rc)--P7(rc) | Old PE library structure |

(Note that the Read2 primer and the i7 Index Read primer are the same location, on opposite strands.)

## CUSTOM SEQUENCING PRIMERS

You may use a custom sequencing primer for the Read1 (side 1) sequence on either the HiSeq or the MiSeq. Please note that acceptable primer Tm varies by instrument - contact the Core for details. IF you need a custom sequencing primer on side 2, then you must use the MiSeq instrument. The only exceptions are 1) if you have 7 lanes, then you can run them together on the HiSeq with the custom primer, or 2) you work with Core lab staff to ensure your primer design does not interfere with the Read2 primer mix.

## SIDE1 vs SIDE2

You must read side1 with a minimum of 25 bases. Even if you don't want the sequence from side1, it is required to set the location coordinates, matrix and phasing information for the run. If there is no data collected for side1, then the side2 read will not be captured. If you want a custom sequence read, e.g. 25 bases on side1 and 125 bases on side2, we can do that on the MiSeq. Please ask. If you want HiSeq data, we can use the closest available run type to your desired length, and bioinformatically trim the Read data. Since in all cases reagents and instrument time are still being used, you should not expect a price break for unlisted read lengths.

## SEQUENCE DIVERSITY

Libraries which begin with a linker, barcode, or other "non-random" sequence will not perform well unless they are base-balanced. This is particularly important on the MiSeq which has only 1 lane. If your sample has the same sequence in the first 6 positions, then we must add a balancer DNA, e.g. PhiX, so that the instrument can be calibrated to capture each base type throughout the sequencing run. If your sample has barcodes in these first 6 positions, and they are base-balanced, it should be fine. If they are not balanced, then we will need to add a balancer DNA. Base-balanced sequences have 1 of each base at each position. For barcodes, we can go with 3 of 4 bases at each of the first 3 positions if each of the four bases is present at least twice.

For Example: (AGC) (CTG) (ATC)

If you have any questions, please ask. If you are submitting libraries with a custom sequencing primer, please include a diagram and the results of any topo cloning or other documentation. The more we know the better we can help you.