

Traditional HPLC is Incapable of Reducing Cross Contamination of Custom Next-Generation Sequencing Adapters to Acceptable Levels

The development of next-generation massively parallel sequencing technologies (NGS), including the Roche 454™ Genome Sequencer FLX Instrument, the Illumina Genome Analyzer, Life Technology's Ion Torrent™ Personal Genome Machine, and the Pacific Biosciences RS have revolutionized genomic and genetic research, significantly improved sequencing throughput, reduced costs for data production, and advanced research from weeks to hours.

Though there are different kinds of NGS technologies including sequencing by synthesis, sequencing by ligation, sequencing by hybridization, single molecule DNA sequencing, and nanopore sequencing, there are many similarities among the sequencing platforms. These can be grouped into library construction, library amplification, parallel sequencing, and data analysis. Synthetic oligonucleotides are used in each of these steps except in the data analysis stage. Each oligonucleotide type has a different set of manufacturing specifications needed to produce consistent results regardless of the sequencing instrument. Based on our experience and close collaboration with customers, we have found that the adapter sequences, which are ligated to sample DNA fragments and are used as both a way to bind fragments to the sequencing media and as an identification marker during multiplexing, carry the most stringent of specifications. The two most important aspects of adapter manufacturing are purity and levels of cross contamination.

In this note, we will focus on how traditional HPLC methods could lead to unwanted (yet undetected) levels of cross contamination, leading to false positives in multiplexing experiments—an expensive problem in terms of time and money.

Before one can determine if a manufacturing process meets minimum criteria for cross contamination, the proper analytical methods must be chosen. Kircher et. al. described a method for minimizing the degree of false calls coming from cross contamination during multiplexing, which is referred to as "double indexing". (Martin Kircher, Susanna Sawyer, Matthias Meyer. *Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform*, Nucleic Acids Research, 2012, e3-e3, DOI: 10.1093/nar/gkr771). In essence, the addition of indexes on both adapters used in an Illumina sequencer helped in determining which sequences belonged to which original sample of genomic DNA in spite of any pre-existing or caused contamination event where more than the desired index adapter was present. Using this same principle, we designed 12 unique synthetic oligonucleotides and manufactured them in isolation (to prevent contamination) in lieu of genomic



DNA fragments to be sequenced. In this case, these sequences became the second index (each synthetic oligonucleotide had a predetermined base sequence) where the first index was one of Illumina’s pre-designed adapters. A total of 24 indexes were used for this experiment (**Figure 1**).

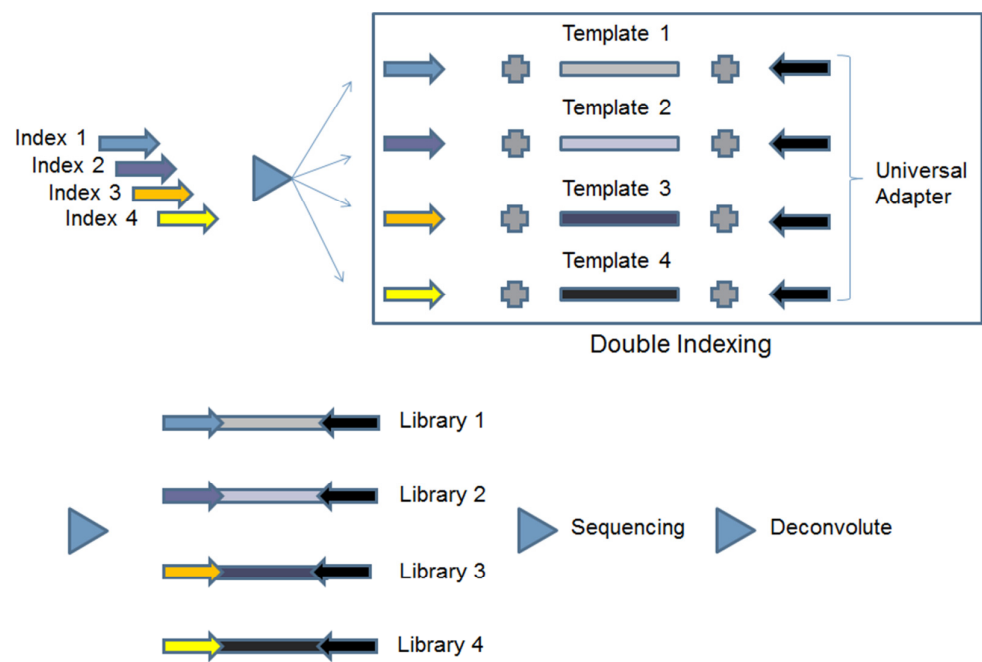


Figure 1. Experimental scheme using double indexing to identify potential cross contamination.

Each index was synthesized and purified using traditional HPLC conditions to recreate what a typical customer would receive from an oligonucleotide manufacturer and then use in a multiplexing experiment. Key to this set up was the fact that the adapter oligonucleotides were purified in sequential order and that the universal adapter was the last in the purification sequence (**Table 1**).

Table 1. Order of the HPLC purification of Illumina Indexes used in this experiment.

Purification Order	ID-1, ID-13, ID-2, ID-14, ID-3, ID-15, ID-4, ID-16, ID-5, ID-17, ID-6, ID-18, ID-7, ID-19, ID-8, ID-20, ID-9, ID-21, ID-10, ID-22, ID-11, ID-23, ID-12, ID-24, Universal
--------------------	--

After ligating purified adapters 1–12 to one of the unique 12 synthetic targets using currently accepted procedures for library construction, the double indexed construct was ready for sequencing. Once the sequencing data were collected and analyzed, the results were summarized on a matrix with the desired index on the horizontal axis and all the other potential contaminants on the vertical axis (**Table 2**).

Table 2. Index abundance for each of the twelve studied Illumina adapters.

Scanned Index	Index 1	Index 2	Index 3	Index 4	Index 5	Index 6	Index 7	Index 8	Index 9	Index 10	Index 11	Index 12
Index 1	89.02%	0.13%	0.02%	0.01%	0.01%	0.00%	0.00%	0.60%	0.00%	0.00%	0.00%	0.01%
Index 2	0.01%	89.36%	0.15%	0.04%	0.01%	0.04%	0.01%	0.01%	0.98%	0.01%	0.01%	0.01%
Index 3	0.00%	0.00%	91.78%	0.17%	0.01%	0.02%	0.04%	0.00%	0.00%	1.43%	0.01%	0.00%
Index 4	0.00%	0.00%	0.00%	91.39%	0.05%	0.04%	0.02%	0.01%	0.00%	0.01%	0.97%	0.00%
Index 5	0.32%	0.02%	0.01%	0.01%	90.51%	0.13%	0.04%	0.02%	0.04%	0.02%	0.02%	0.58%
Index 6	1.21%	0.97%	0.59%	0.92%	0.99%	85.39%	1.30%	0.92%	1.11%	0.80%	0.83%	0.82%
Index 7	0.22%	0.00%	0.00%	0.00%	0.00%	0.00%	86.64%	0.02%	0.01%	0.01%	0.02%	0.01%
Index 8	0.01%	0.00%	0.00%	0.01%	0.01%	0.01%	0.01%	89.76%	0.19%	0.04%	0.02%	0.03%
Index 9	0.05%	0.06%	0.04%	0.05%	0.07%	0.07%	0.05%	0.04%	85.87%	0.28%	0.09%	0.05%
Index 10	0.01%	0.01%	0.01%	0.01%	0.01%	0.03%	0.01%	0.01%	0.02%	83.35%	0.10%	0.02%
Index 11	0.01%	0.02%	0.01%	0.01%	0.02%	0.02%	0.01%	0.01%	0.01%	0.01%	84.73%	0.08%
Index 12	0.13%	0.14%	0.13%	0.11%	0.12%	0.17%	0.13%	0.10%	0.11%	0.11%	0.15%	88.17%
Index 13	0.00%	1.21%	0.03%	0.01%	0.98%	0.01%	0.01%	0.00%	0.08%	0.00%	0.00%	0.10%
Index 14	0.52%	0.53%	1.68%	0.43%	0.43%	4.45%	0.49%	0.41%	0.50%	0.40%	0.73%	0.40%
Index 15	0.00%	0.00%	0.00%	0.90%	0.02%	0.02%	1.53%	0.00%	0.00%	0.00%	0.01%	0.00%
Index 16	0.00%	0.00%	0.00%	0.00%	0.30%	0.03%	0.01%	1.78%	0.00%	0.00%	0.00%	0.01%
Index 17	0.02%	0.02%	0.15%	0.01%	0.01%	1.43%	0.07%	0.02%	3.12%	0.05%	0.03%	0.01%
Index 18	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.46%	0.01%	0.01%	3.03%	0.01%	0.00%
Index 19	0.06%	0.06%	0.06%	0.07%	0.00%	0.09%	0.00%	0.09%	0.01%	0.02%	3.11%	0.10%
Index 20	0.01%	0.01%	0.01%	0.00%	0.01%	0.01%	0.02%	0.01%	1.16%	0.08%	0.03%	2.76%
Index 21	2.94%	2.62%	2.52%	2.40%	4.14%	3.69%	2.76%	2.48%	2.78%	3.63%	3.34%	2.29%
Index 22	0.01%	0.02%	0.01%	0.01%	0.02%	0.02%	0.02%	0.01%	0.01%	0.01%	0.85%	0.05%
Index 23	0.15%	0.08%	0.09%	0.08%	0.04%	0.10%	0.09%	0.09%	0.09%	0.26%	0.19%	0.68%
Index 24	1.14%	1.10%	0.96%	1.01%	1.10%	1.46%	0.99%	1.04%	1.06%	0.96%	1.35%	0.84%
Unmapped	4.14%	3.65%	1.72%	2.34%	1.16%	2.75%	4.29%	2.56%	2.82%	5.47%	3.40%	2.96%

Based on these experimental conditions, the analyzed data revealed that under traditional HPLC conditions, sequential purification of adapter oligonucleotides led to cross contamination from the previously purified adapter, which ranged from 0.30–1.68% relative to the total number of sequencing reads (excluding the contribution of other contamination events). Furthermore, potential contamination from an adapter purified prior to the one immediately before the desired sequence could also be detected, though at much lower levels (0.02–0.17%). Additional care should be taken when also purifying the universal adapter as any contamination present on this oligonucleotide will carry over to all other indexes in the multiplex experiment. This experimental set up also revealed that there are other steps upstream and downstream of the purification process that could lead to cross contamination. The latter is evident in the presence of contamination throughout all indexes of Index 6, 12, 14, 21, and 24 (**Table 2**).

The modification of the HPLC conditions as well as downstream and upstream steps to account for these observations helped produce adapter oligonucleotides which showed cross contamination levels below 0.3%. Since other factors can contribute to the observed cross-contamination percentage when the oligonucleotides are used for NGS sequencing that do not have to do with the manufacturing of the oligonucleotides (see reference before), it is paramount to have processes that minimize such events. Our engineers and scientists have used the above information to develop a set of manufacturing conditions to deliver oligonucleotides used in multiplexing experiments that can accommodate the different NGS needs of customers.