

Illumina Unamplified Indexed Library Construction: An Automated Approach

Christopher A. Hack¹, Alex Sczyrba¹, Jan-Fang Cheng¹

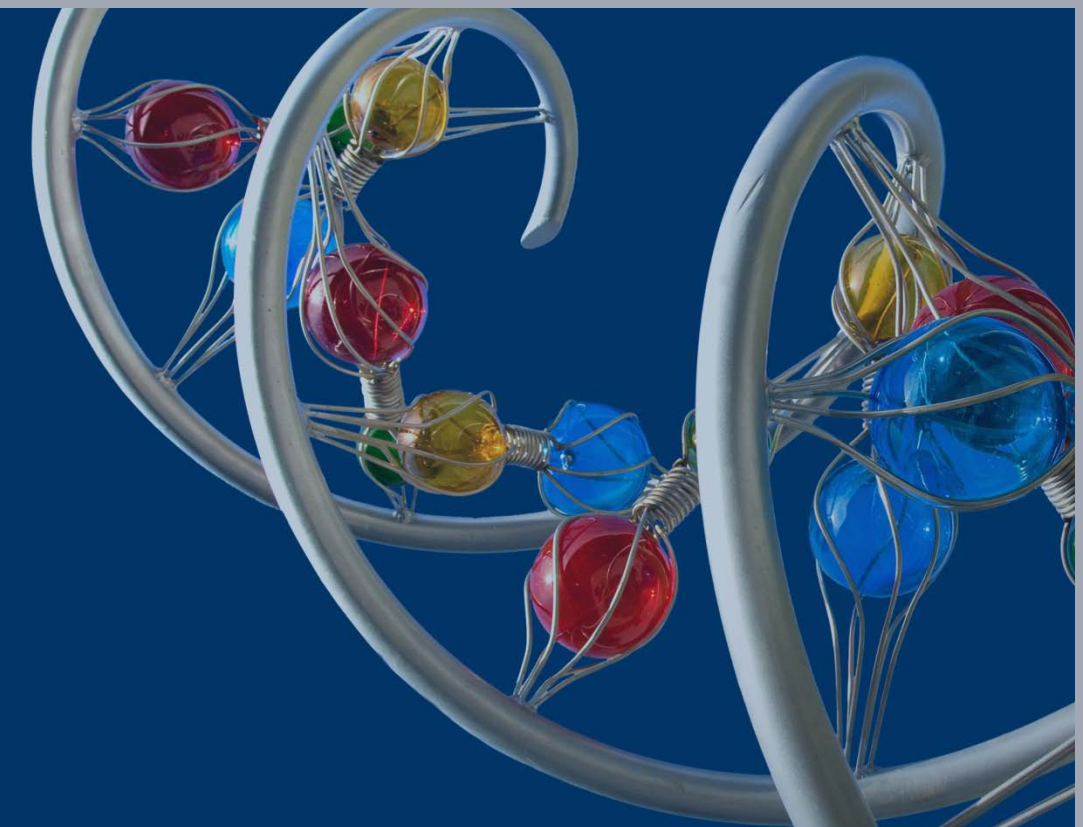
¹Lawrence Berkeley National Laboratory

March 2011

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.

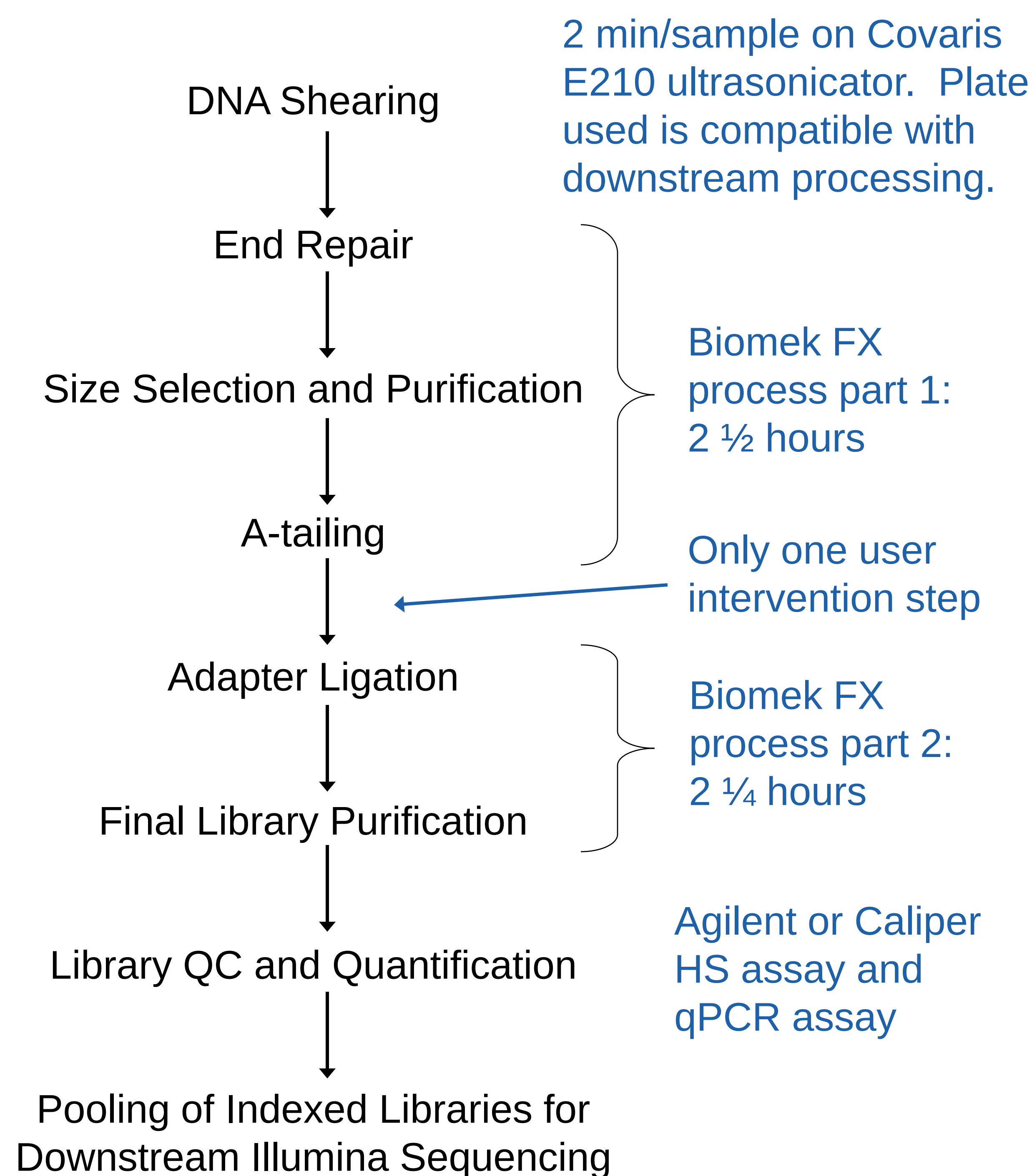


Introduction

Manual library construction is a limiting factor in Illumina sequencing. Constructing libraries by hand is costly, time-consuming, low-throughput, and ergonomically hazardous, and constructing multiple libraries introduces risk of library failure due to pipetting errors. The ability to construct multiple libraries simultaneously in automated fashion represents significant cost and time savings.

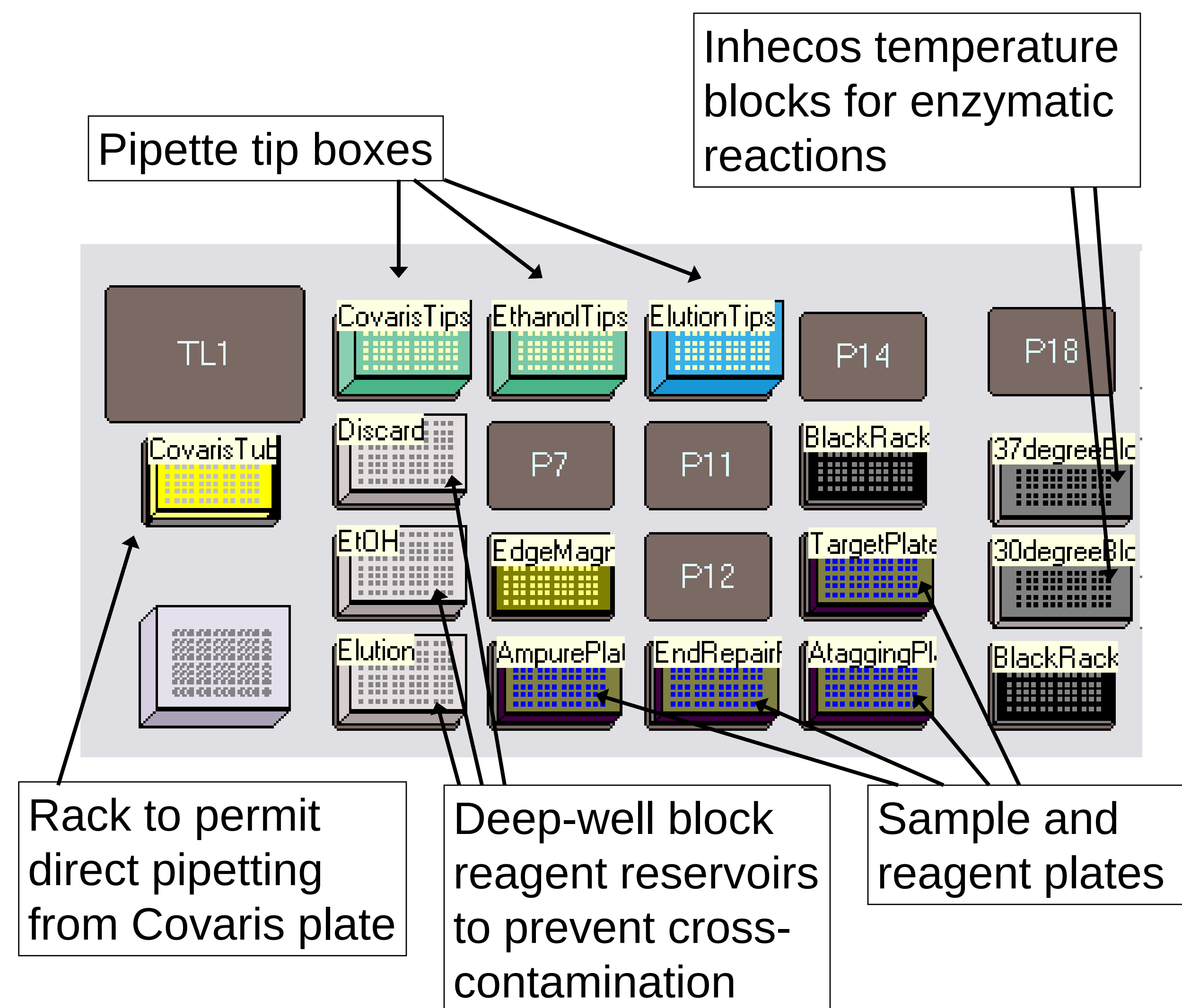
Here we present a strategy to construct up to 96 unamplified indexed libraries using Illumina TruSeq reagents and a Biomek FX robotic platform. We also present data to indicate that this library construction method has little or no risk of cross-contamination between samples.

Workflow



Following DNA shearing, a single operator can construct 96 libraries and set up QC in a single day.

Instrument Layout

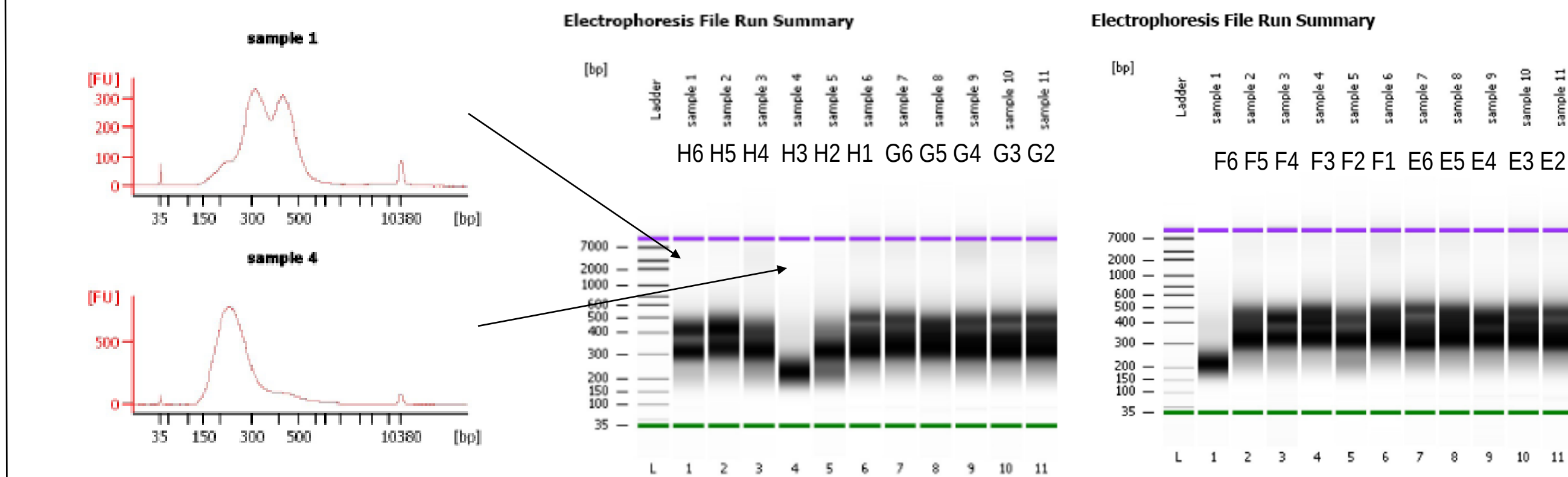


Proof of Principle Experiment

24 unamplified indexed TruSeq libraries were constructed using this method and subsequently pooled and sequenced in a 1x36 run on an Illumina GAii instrument. Sequence data was analyzed for potential cross-contamination between adjacent libraries.

Column	1	2	3	4	5	6
Row E	E.coli K12 Adapter #9	E.coli K12 Adapter #8	E.coli K12 Adapter #7	E.coli K12 Adapter #12	E.coli K12 Adapter #11	E.coli K12 Adapter #10
Row F	E.coli K12 Adapter #3	Saccharomonospora azurea Adapter #2	E.coli K12 Adapter #1	Actinopolyspora iraqiensis Adapter #6	E.coli K12 Adapter #5	Saccharomonospora saliphila Adapter #4
Row G	E.coli K12 Adapter #12	E.coli K12 Adapter #11	E.coli K12 Adapter #10	E.coli K12 Adapter #9	E.coli K12 Adapter #8	E.coli K12 Adapter #7
Row H	E.coli K12 Adapter #6	Cycloclostridium pugetii Adapter #5	E.coli K12 Adapter #4	Acetobacterium sp. WL Adapter #3	E.coli K12 Adapter #2	Sporomusa sp. KB1 Adapter #1

Library QC Data



Top: Successful library: electropherogram shows mono- and duo-adaptor peaks

Bottom: Failed library: electropherogram shows no adapter ligation

Agilent HS data for libraries in rows H and G

Agilent HS data for libraries in rows F and E

Both libraries constructed with Adapter #4 failed, leading us to conclude that Adapter #4 in the kit lot used was faulty, rather than the method failing.

Pool #1: Libraries in rows H and G, except failed library H3 was not included.
Pool #2: Libraries in rows F and E, except failed library F6 was not included.

Sequence data analysis

From each pool, reads were sorted by index sequence, then mapped against the reference genomes of *S. azurea* (library F2), *A. iraqiensis* (library F4), and *E. coli* (control libraries). The percentage of reads mapping to each reference genome are shown in the table below:

	1	2	3	4	5	6
E	0.09%	0.06%	0.08%	0.06%	0.07%	0.06%
E	0.23%	0.19%	0.21%	0.19%	0.21%	0.20%
E	95.54%	95.74%	95.41%	95.37%	95.88%	95.55%
F	0.05%	85.40%	0.07%	4.64%	0.06%	0.07%
F	0.19%	4.19%	0.20%	79.97%	0.19%	0.20%
F	95.78%	1.27%	95.20%	1.96%	95.38%	95.47%
G	0.05%	0.05%	0.05%	0.05%	0.05%	0.05%
G	0.19%	0.20%	0.19%	0.19%	0.19%	0.19%
G	95.31%	95.85%	95.41%	95.45%	95.67%	95.35%
H	0.05%	0.01%	0.05%	0.03%	0.05%	0.05%
H	0.19%	0.02%	0.19%	0.03%	0.19%	0.04%
H	95.58%	0.35%	95.41%	0.17%	95.43%	0.43%

For each index there are a handful of reads that mapped to the “wrong” reference genome. When the mapping is redone excluding reads from ribosomal DNA, the instances of “wrong” mapping are not seen. This indicates that the reads mapping to unexpected reference genomes are not a result of cross-contamination during library construction, but rather just representative of highly conserved regions of ribosomal DNA.

Reads corresponding to Index #4 were observed in both pools, despite the fact that no library constructed with Adapter #4 was included in either pool. We hypothesize that this is the result of contamination of kit reagents, and was not introduced during library construction.

Conclusion

We have successfully demonstrated the ability to construct up to 96 Illumina unamplified indexed libraries in high-throughput fashion, and that the risk of cross-contamination in the library construction process is minimal. We have high hopes of being able to scale up this process for use in a production sequencing environment.