From sample to fastq

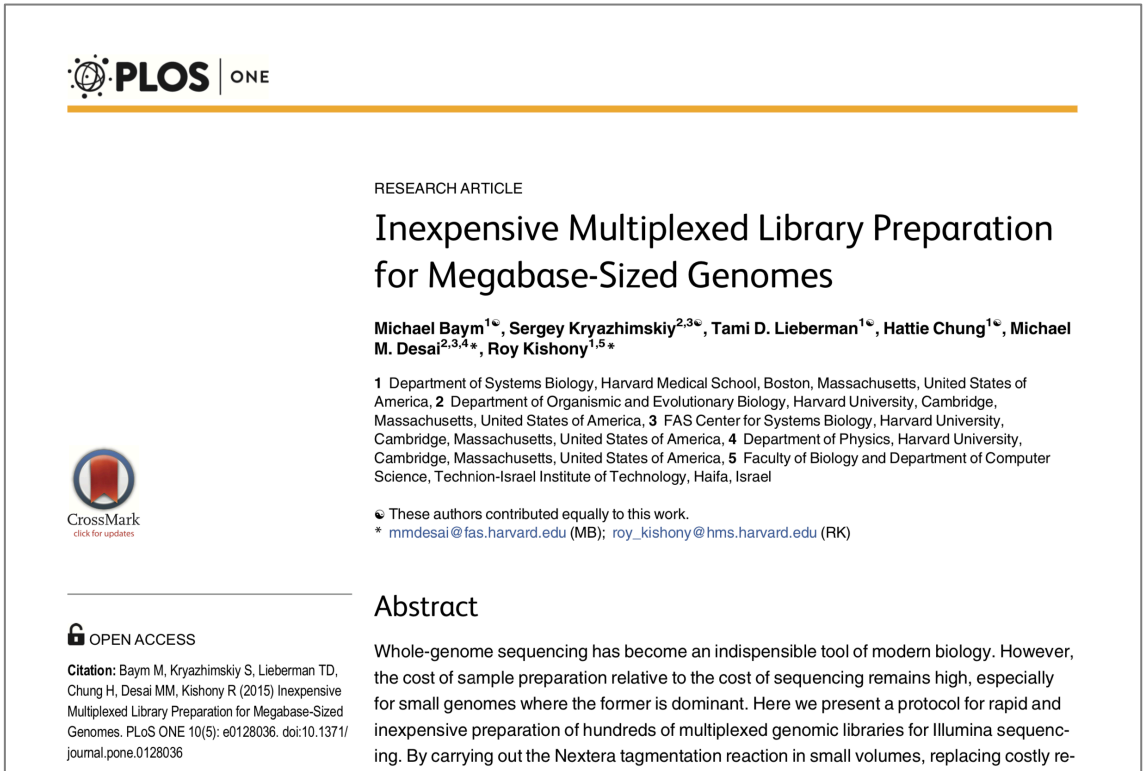
Outline

- Brief overview of library preparation procedure
- Sequencing costs
- Estimate cost for your own experiment

Requirements for library preparation protocol

- To prepare libraries for hundreds of samples, we need a protocol that is
  - Cheap
  - Efficient
  - Reliable
- Sometimes robustness to sample degradation is also important

One example of a library preparation technique



PLOS | ONE

RESEARCH ARTICLE

Inexpensive Multiplexed Library Preparation for Megabase-Sized Genomes

Michael Baym[1☯], Sergey Kryazhimskiy[2,3☯], Tami D. Lieberman[1☯], Hattie Chung[1☯], Michael M. Desai[2,3,4*], Roy Kishony[1,5*]

1 Department of Systems Biology, Harvard Medical School, Boston, Massachusetts, United States of America, **2** Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, United States of America, **3** FAS Center for Systems Biology, Harvard University, Cambridge, Massachusetts, United States of America, **4** Department of Physics, Harvard University, Cambridge, Massachusetts, United States of America, **5** Faculty of Biology and Department of Computer Science, Technion-Israel Institute of Technology, Haifa, Israel

☯ These authors contributed equally to this work.
* mmdesai@fas.harvard.edu (MB); roy_kishony@hms.harvard.edu (RK)

CrossMark
click for updates

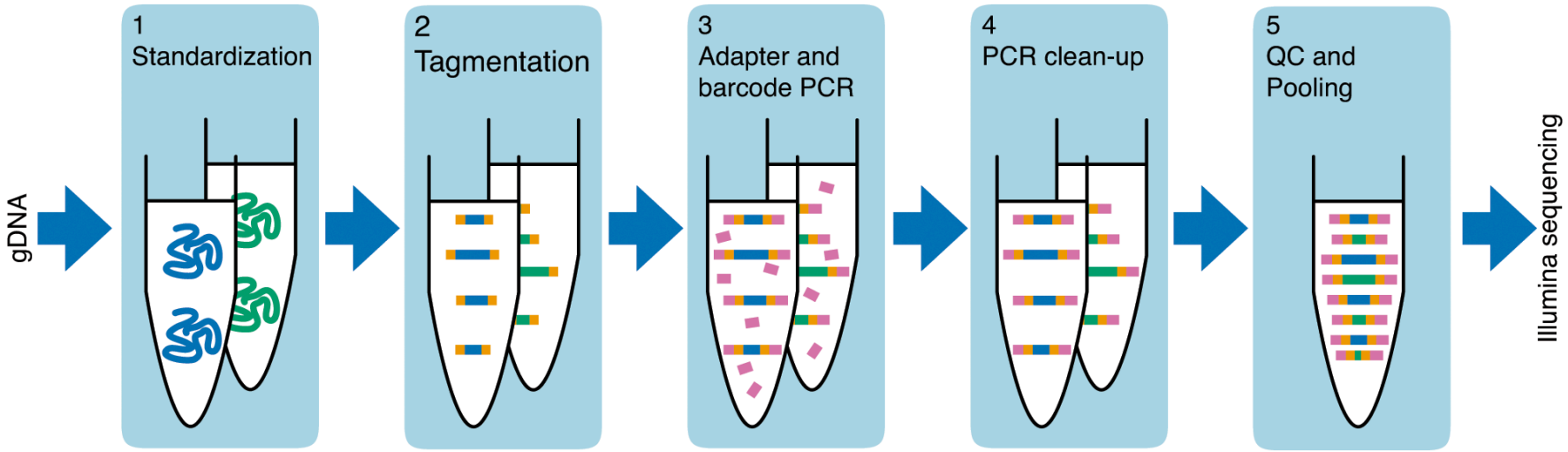
OPEN ACCESS

Citation: Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony R (2015) Inexpensive Multiplexed Library Preparation for Megabase-Sized Genomes. PLoS ONE 10(5): e0128036. doi:10.1371/journal.pone.0128036

Abstract

Whole-genome sequencing has become an indispensible tool of modern biology. However, the cost of sample preparation relative to the cost of sequencing remains high, especially for small genomes where the former is dominant. Here we present a protocol for rapid and inexpensive preparation of hundreds of multiplexed genomic libraries for Illumina sequencing. By carrying out the Nextera tagmentation reaction in small volumes, replacing costly re-

Library preparation protocol



Transposome with adapters combined with template DNA

Tagmentation to fragment and add adapters

Limited-cycle PCR to add index adapter sequences

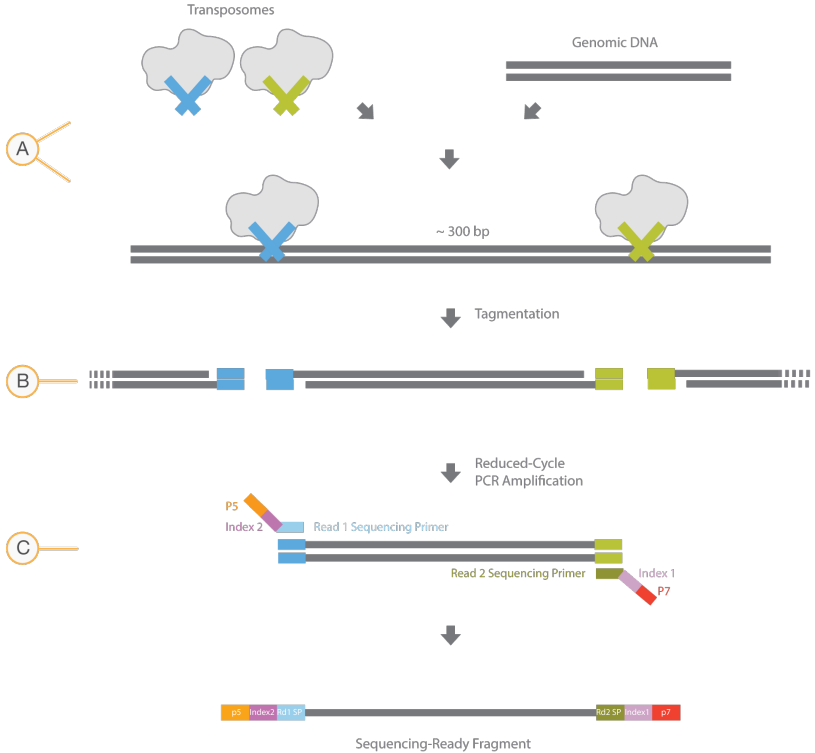


Image from the Nextera XT DNA Library Prep Kit Reference Guide (© 2019 Illumina, Inc.)

Sequencing-Ready Fragment

Index 2 Primer

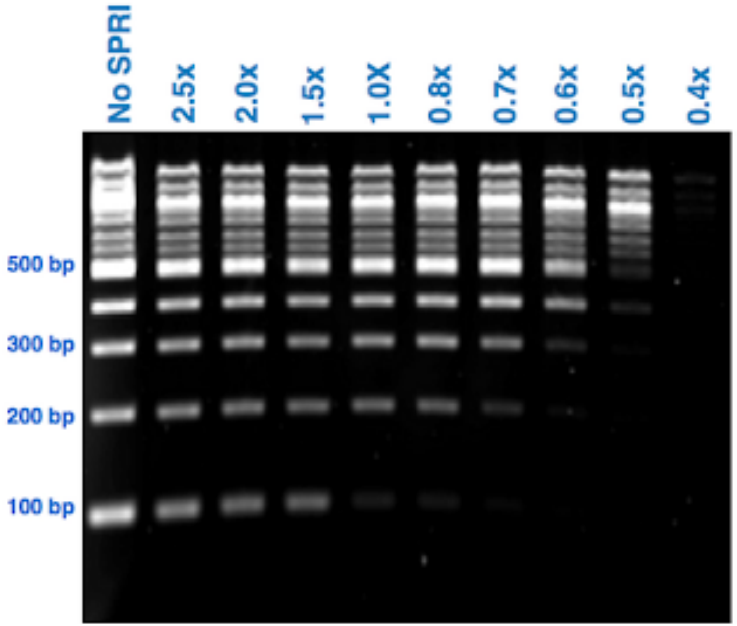
Index 1 Primer

P5 – complementary to Illumina flow cell oligo
Indexing sequence 2
Read 1 Sequencing Primer

Read 2 Sequencing Primer
Indexing sequence 1
P7 – complementary to Illumina flow cell oligo

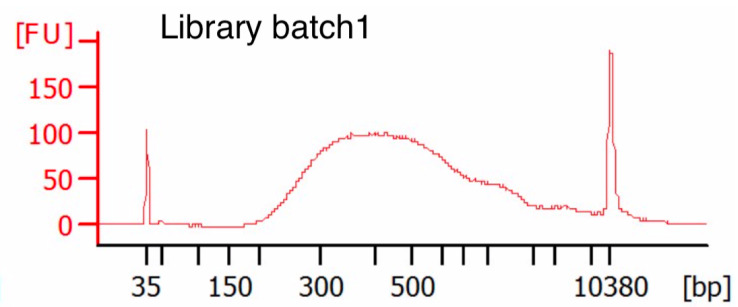
Image adapted from the Nextera XT DNA Library Prep Kit Reference Guide (© 2017 Illumina, Inc.)

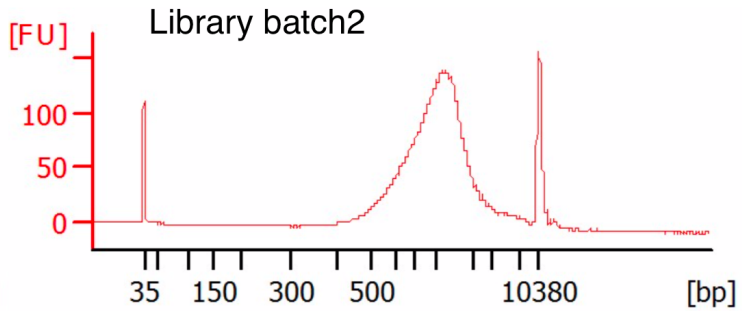
Size selection with Ampure beads

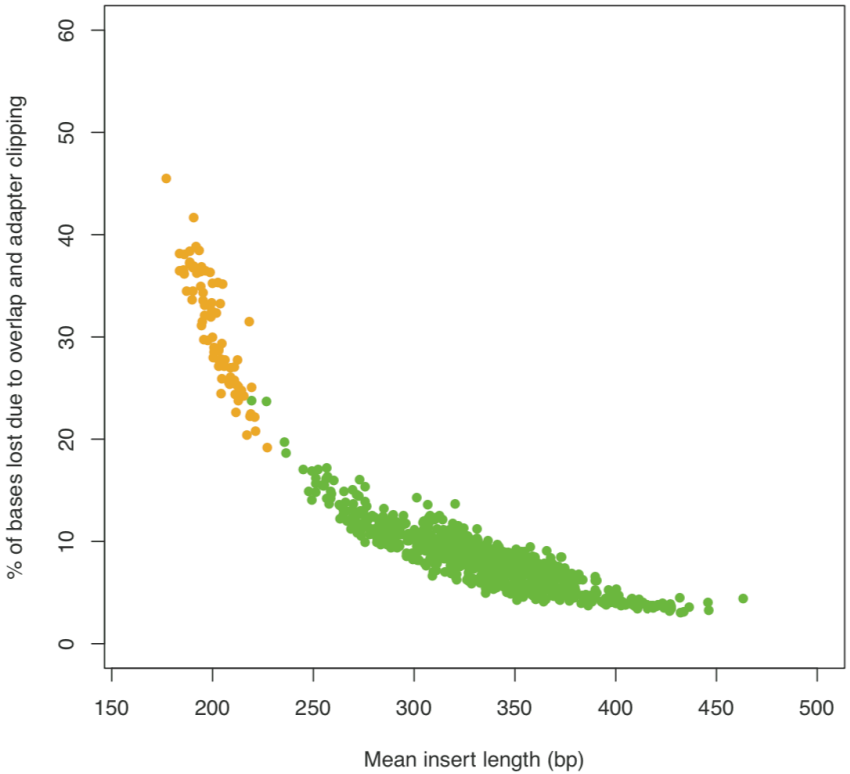


http://enseqlopedia.com/2012/04/how-do-spri-beads-work/

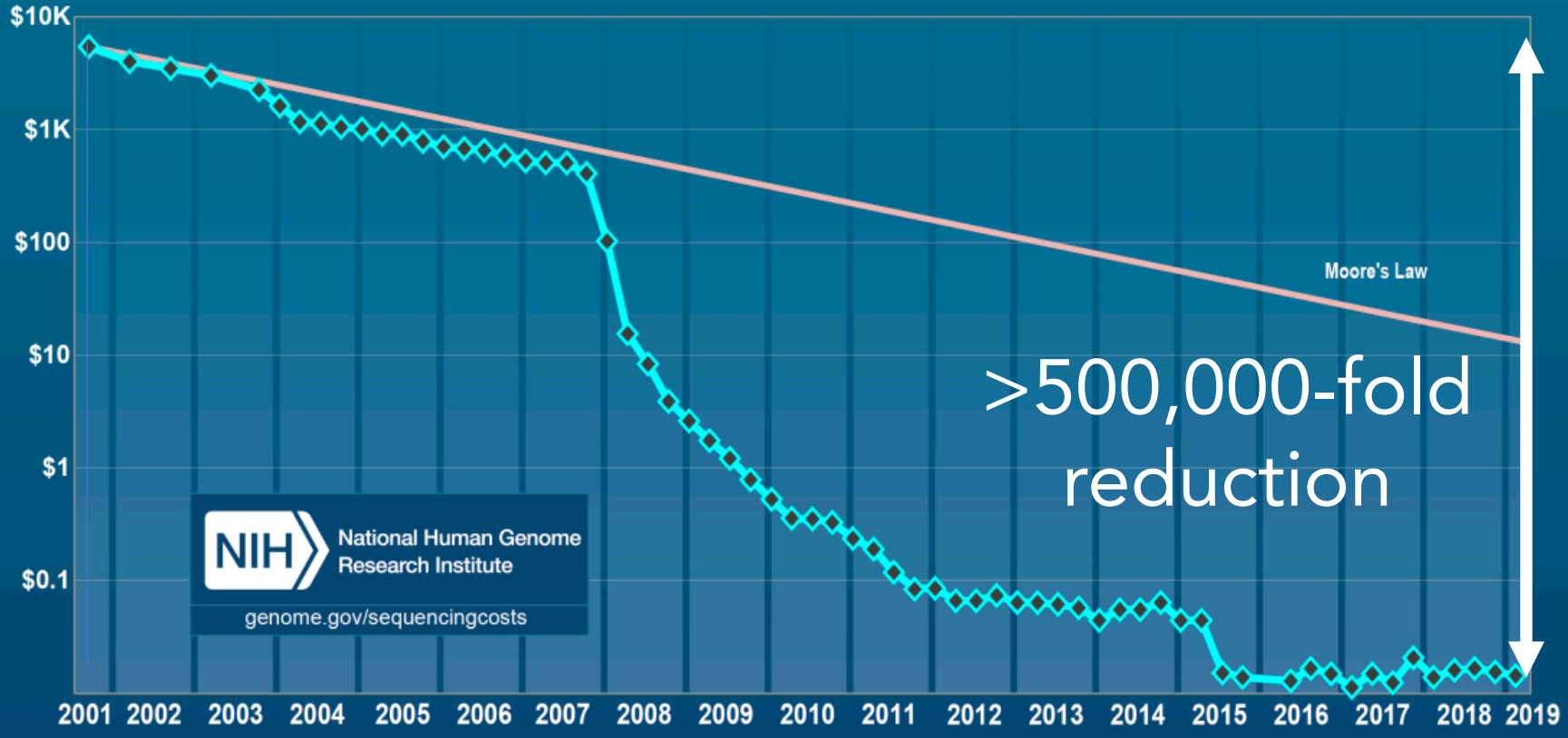
Two examples of our library pools






% of bases lost due to overlap and adapter clipping
0
10
20
30
40
50
60
150
200
250
300
350
400
450
500
Mean insert length (bp)

Cost per Raw Megabase of DNA Sequence



What is the current price for 2x sequencing of an Atlantic silverside?

Genome size ~650 Mb

- $12
- $31
- $48
- $72
- $213

What is the current price for 2x sequencing of an Atlantic silverside?

Genome size ~650 Mb

- $12
- $31
- $48
- $72
- $213

Example costs for other genome sizes

Library preparation and sequencing to 2x genome coverage*

Genome size (Mb)	Cost (USD)
500	28
750	34
1,000	40
3,000	87

*Cost estimates do not include labor and assume that samples are sequenced efficiently on 110 Gb HiSeq X Ten lanes (assumed costs: USD 1,300 per lane and USD 8 per library)

Example costs for other genome sizes

Library preparation and sequencing to 2x coverage*

Genome size (Mb)	Cost (USD)
500	28
750	34
1,000	40
3,000	87

Compare to:

$30 per sample for RADseq
$15 per sample for RADcapture

Meek and Larson. 2019. Mol Ecol Res

*Cost estimates do not include labor and assume that samples are sequenced efficiently on 110 Gb HiSeq X Ten lanes (assumed costs: USD 1,300 per lane and USD 8 per library)

Exercise – how much will your experiment cost?

- Assumed costs:
  - Library preparation: $8 per sample
  - Sequencing: $1,300 for 110 Gb
  - Target coverage per sample: Expect to loose at least 30-50% of your data in filtering

Exercise – how much will your experiment cost?

- Assumed costs:
  - Library preparation: $8 per sample
  - Sequencing: $1,300 for 110 Gb
  - Target coverage per sample: Expect to loose at least 30-50% of your data in filtering

- Example: I would like to have 1x coverage for downstream analysis for 40 individuals from each of 5 populations (200 individuals total) of my favorite animal with a genome size of ~800 Mb

- Calculation: I will target 2x coverage raw sequencing. This means

  2 * 800 Mb/individual * 200 individuals = 320,000 Mb (320 Gb) [I can fit this in three 110 Gb lanes]

  My total cost is thus (3 lanes * $1,300 per lane) + 200 libraries * $8 per library) = **$5,500**