Sample Preparation

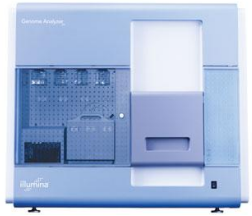
Module 1: Overview

BROAD INSTITUTE illumina

Sequencing Workflow

Sample Preparation

Cluster Generation

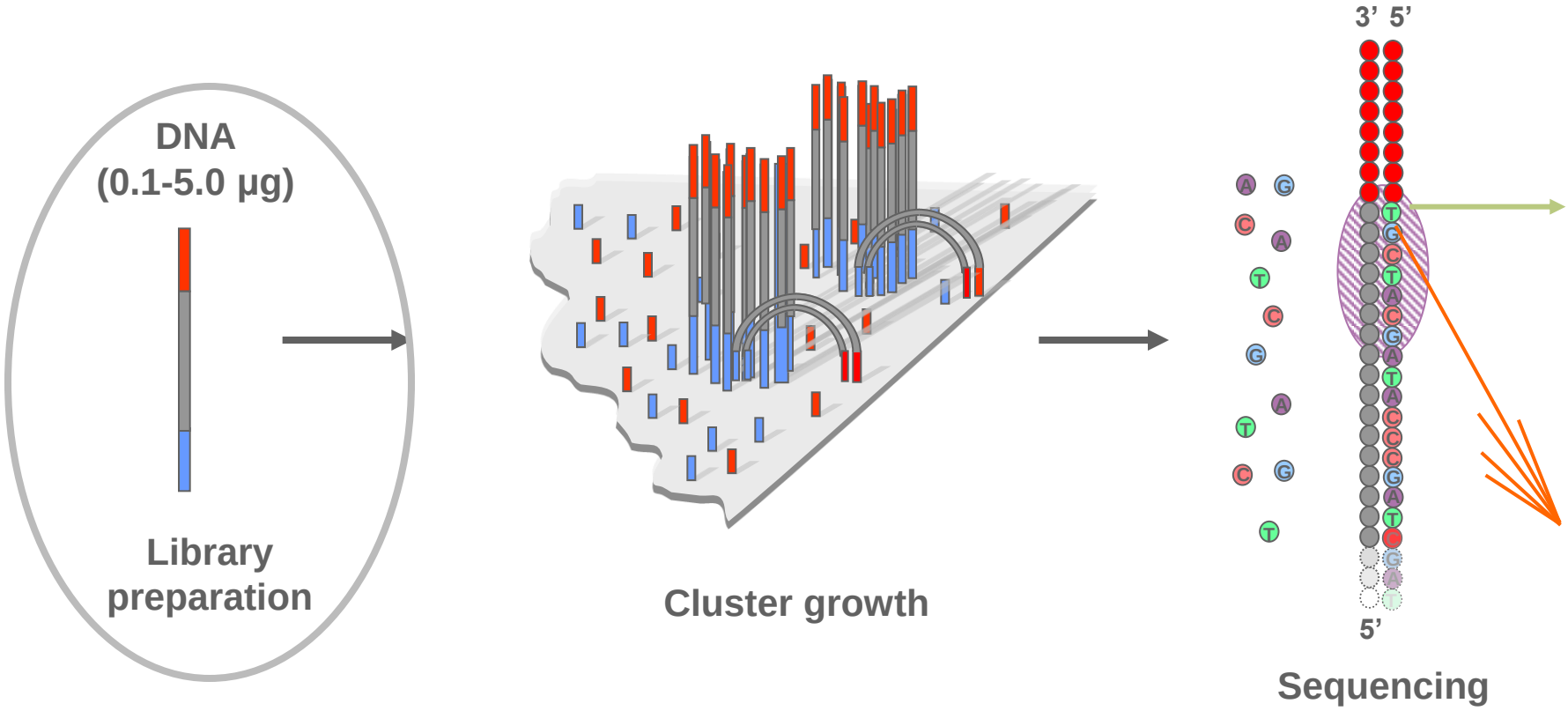


Sequencing

Data Analysis

BROAD INSTITUTE illumina

Library Preparation



BROAD INSTITUTE illumina®

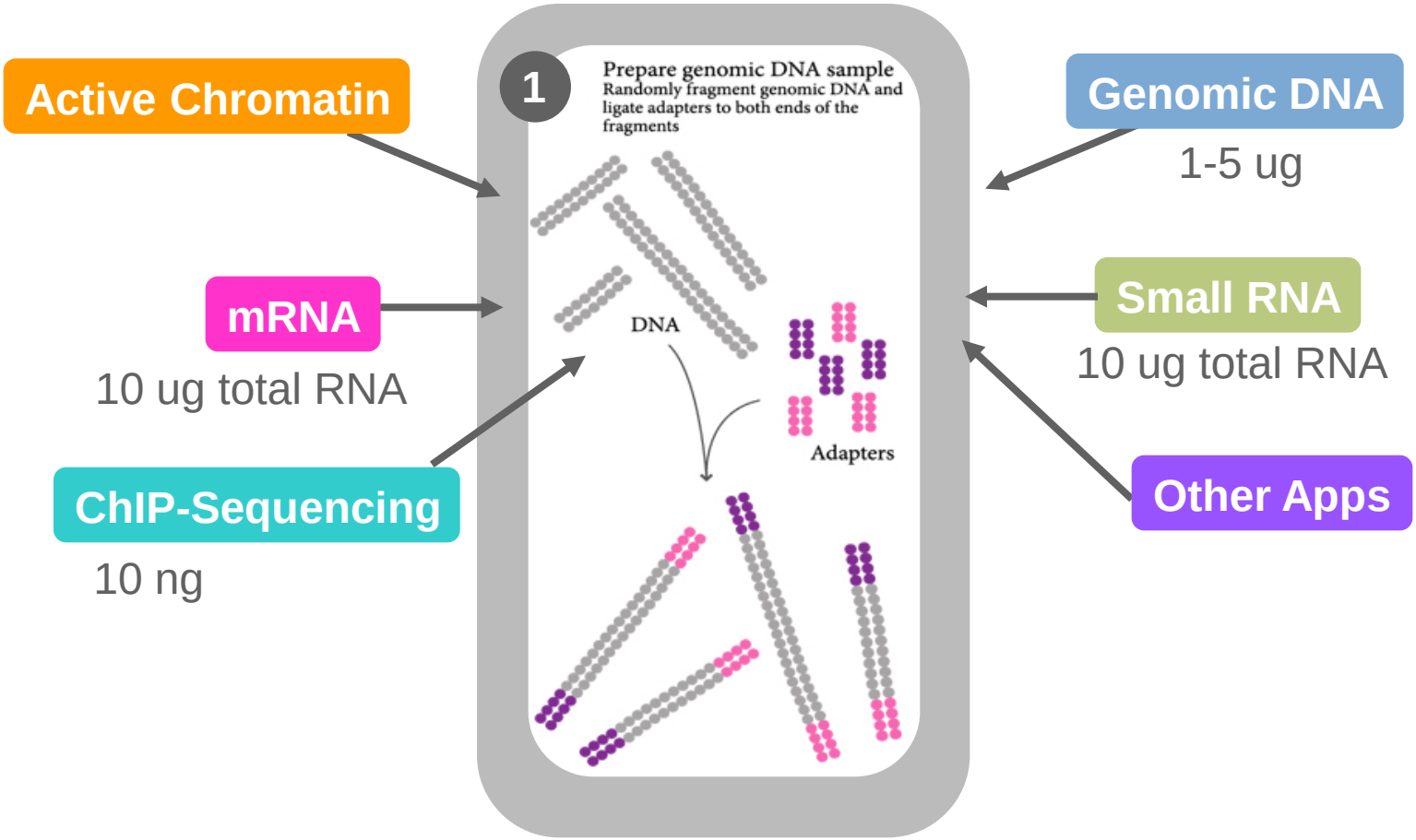
Library Preparation

Sample Preparation

- Prepares sample nucleic acid for sequencing
  - Fragmenting
  - Generates double-stranded DNA (if necessary)
  - Flanks with Illumina adapters
- All preparation ends with the same general template structure
  - Double-stranded DNA flanked by adapters
  - Variables include
    - Insert Size
    - Adaptor type
    - Index

BROAD INSTITUTE illumina

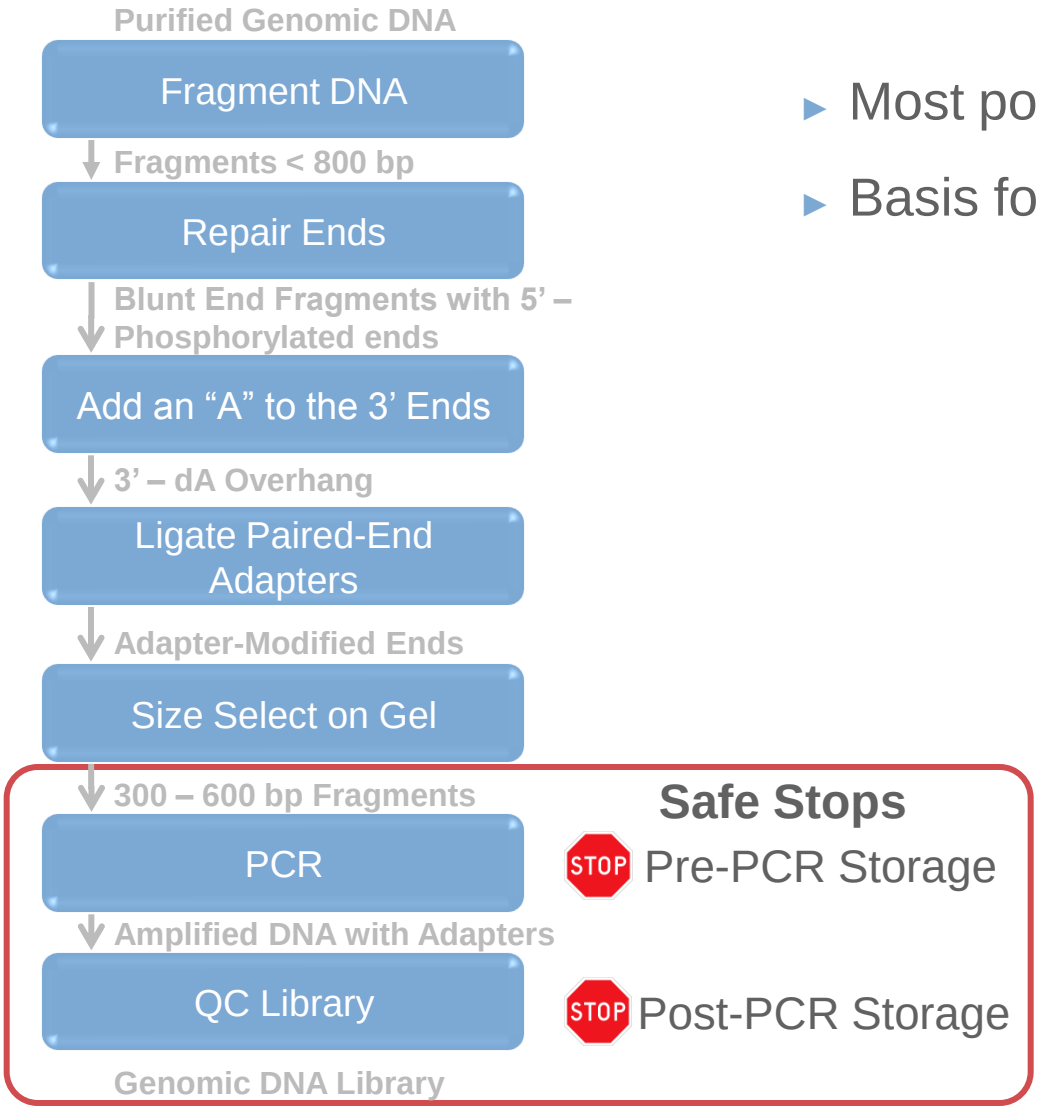
Library Preparation



Library preparation:
Attach adapters to both ends of fragmented nucleic acid

BROAD INSTITUTE illumina

Genomic DNA Protocol Overview

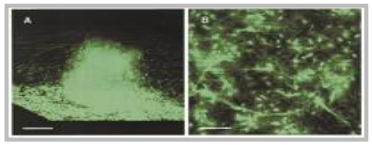


- Most popular protocol
- Basis for all other protocols

BROAD INSTITUTE illumina

Starting Material Input

- Starting Materials
  - Quantify
  - Quality Control
- Methods for Quantification:
- 




  NanoDrop* PicoGreen® * Qubit™

- Methods for Quality Control
  - Visualization on gel
  - Agilent Bioanalyzer* trace especially useful for RNA (RIN)

***Methods utilized at the Broad**

BROAD INSTITUTE illumina®

Step 1: Fragment DNA

Purified Genomic DNA

Fragment DNA

Fragments < 800 bp

Repair Ends

Blunt End Fragments with 5' – Phosphorylated ends

Add an "A" to the 3' Ends

3' – dA Overhang

Ligate Paired-End Adapters

Adapter-Modified Ends

Size Select on Gel

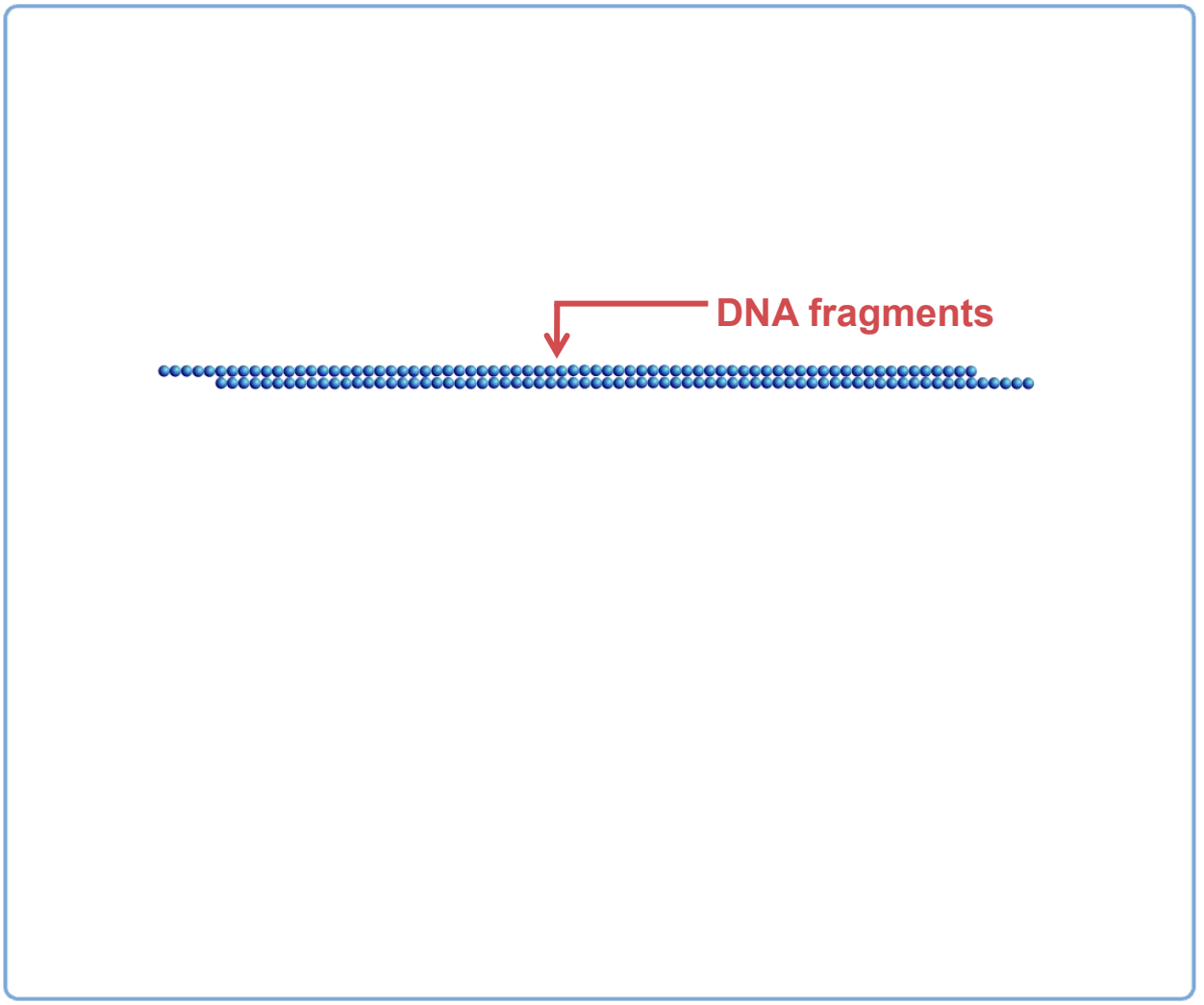
300 – 600 bp Fragments

PCR

Amplified DNA with Adapters

QC Library

Genomic DNA Library



BROAD INSTITUTE illumina

Step 1: Fragment DNA

Purified Genomic DNA

Fragment DNA

Fragments < 800 bp

Repair Ends

Blunt End Fragments with 5' – Phosphorylated ends

Add an "A" to the 3' Ends

3' – dA Overhang

Ligate Paired-End Adapters

Adapter-Modified Ends

Size Select on Gel

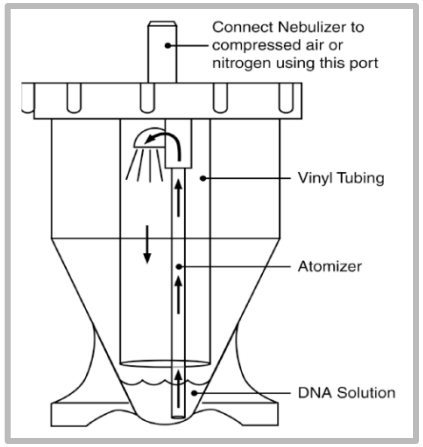
300 – 600 bp Fragments

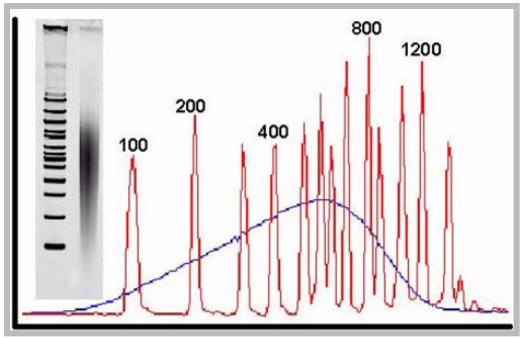
PCR

Amplified DNA with Adapters

QC Library

Genomic DNA Library

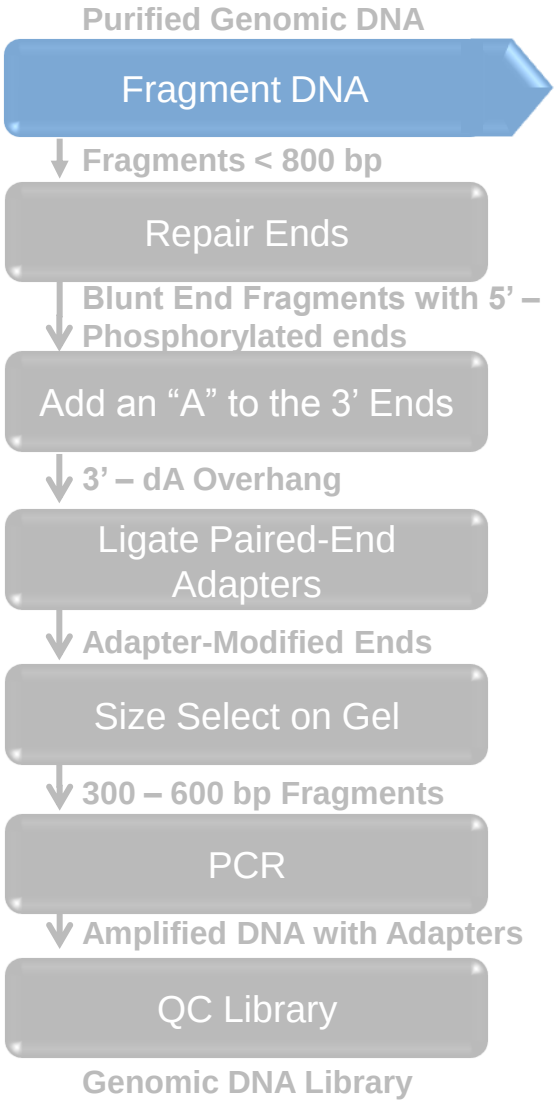




- Nebulizer
  - Very inexpensive
  - Works well for 1-5 µg starting input
- Other methods available
  - Covaris™
  - Sonication
  - HydroShear®
  - Enzymatic
  - Others

BROAD INSTITUTE illumina®

Step 1: Fragment DNA



	Nebulizer	Covaris™	HydroShear®	Sonication	Chemical	Enzymatic
DNA	✓	✓	✓	✓		
mRNA					✓	
DGE						✓
small RNA						
CHIP-Seq		✓		✓		
Mate Pair	✓		✓			

- Fragmentation methods have benefits and limitations
- Can be application-specific
- Depends on desired end-product

BROAD INSTITUTE illumina®

Step 2: Repair Ends

Purified Genomic DNA

Fragment DNA

Fragments < 800 bp

Repair Ends

Blunt End Fragments with 5' – Phosphorylated ends

Add an "A" to the 3' Ends

3' – dA Overhang

Ligate Paired-End Adapters

Adapter-Modified Ends

Size Select on Gel

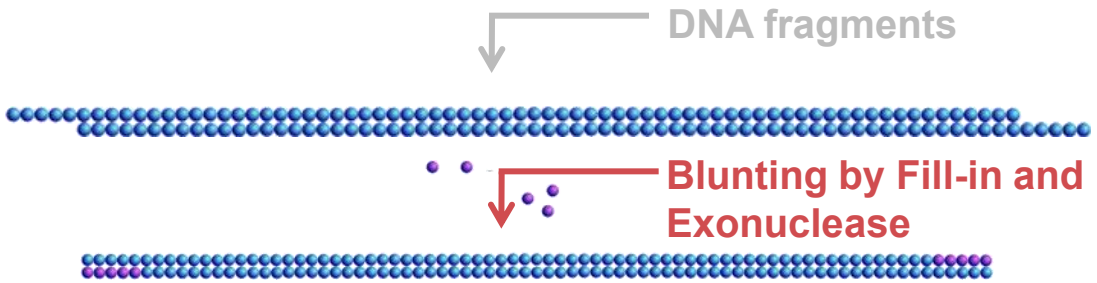
300 – 600 bp Fragments

PCR

Amplified DNA with Adapters

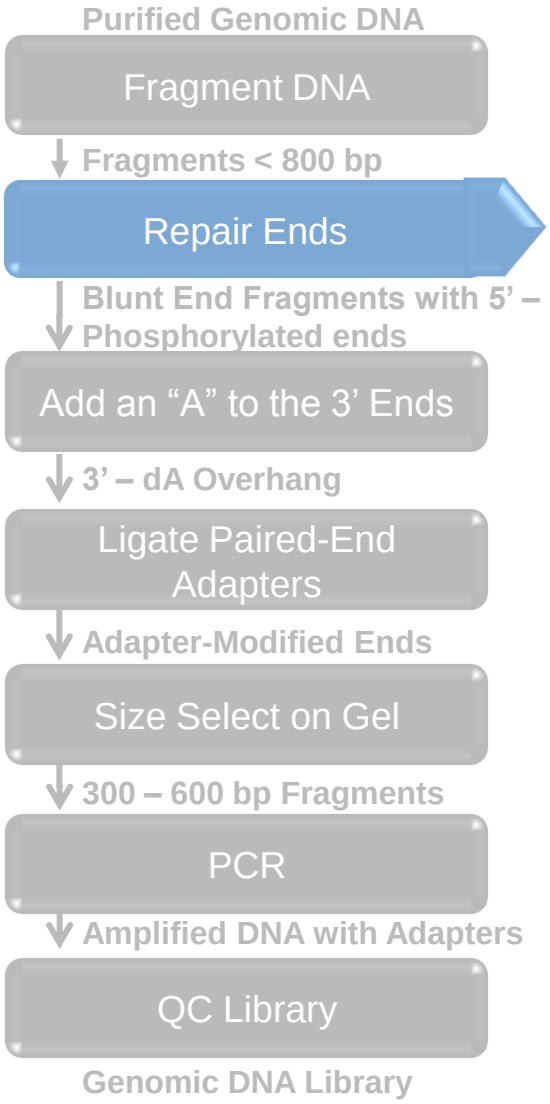
QC Library

Genomic DNA Library



BROAD INSTITUTE illumina

Step 2: Repair Ends



- Converts overhangs
  - Blunt ends
  - Phosphorylates 5' End
- Reagents:
  - dNTP, T4 DNA pol, Klenow
  - Kinase/ATP (T4 PNK)
- Simple enzymatic reaction
- 30 minutes incubation

BROAD INSTITUTE illumina

Step 3: Add “A” Overhang

Purified Genomic DNA

Fragment DNA

Fragments < 800 bp

Repair Ends

Blunt End Fragments with 5’ – Phosphorylated ends

Add an “A” to the 3’ Ends

3’ – dA Overhang

Ligate Paired-End Adapters

Adapter-Modified Ends

Size Select on Gel

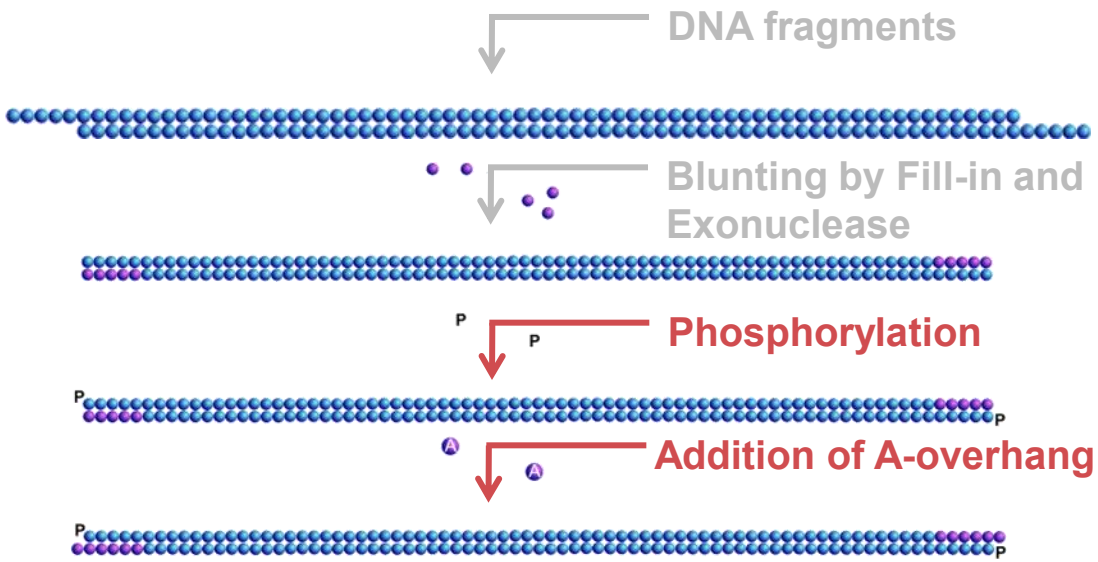
300 – 600 bp Fragments

PCR

Amplified DNA with Adapters

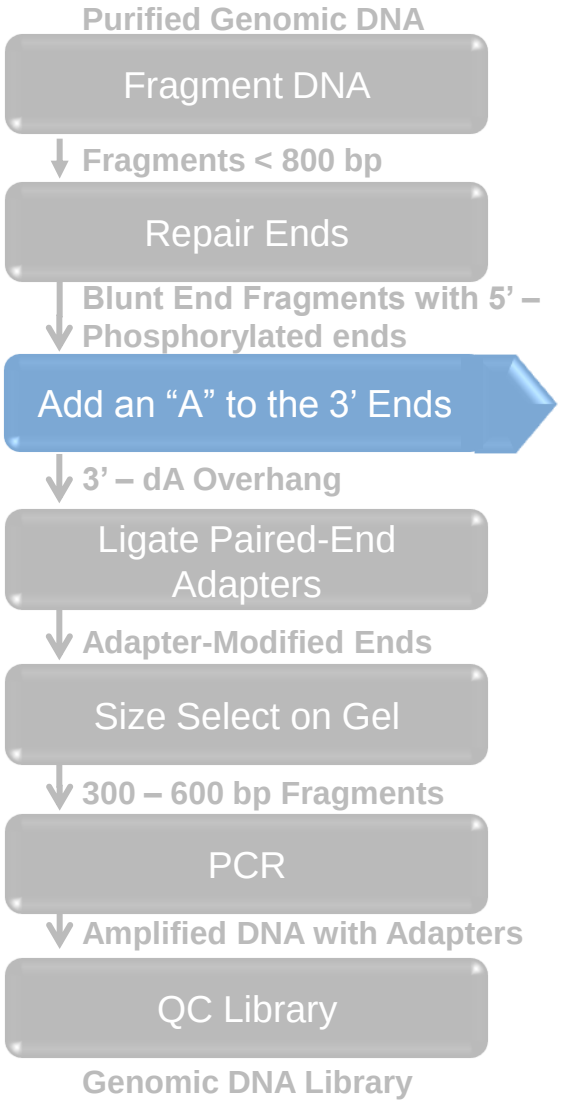
QC Library

Genomic DNA Library



BROAD INSTITUTE illumina

Step 3: Add “A” Overhang



- Adds
  - ‘A’ base to the 3' end of the blunt phosphorylated DNA fragments
- Prevents
  - Formation of adapters dimers
  - Concatemers
- Reagents
  - 1 mM dATP, Klenow exo (3' to 5' exo minus)

BROAD INSTITUTE illumina

Step 4: Ligate Adapters

Purified Genomic DNA

Fragment DNA

Fragments < 800 bp

Repair Ends

Blunt End Fragments with 5' – Phosphorylated ends

Add an "A" to the 3' Ends

3' – dA Overhang

Ligate Paired-End Adapters

Adapter-Modified Ends

Size Select on Gel

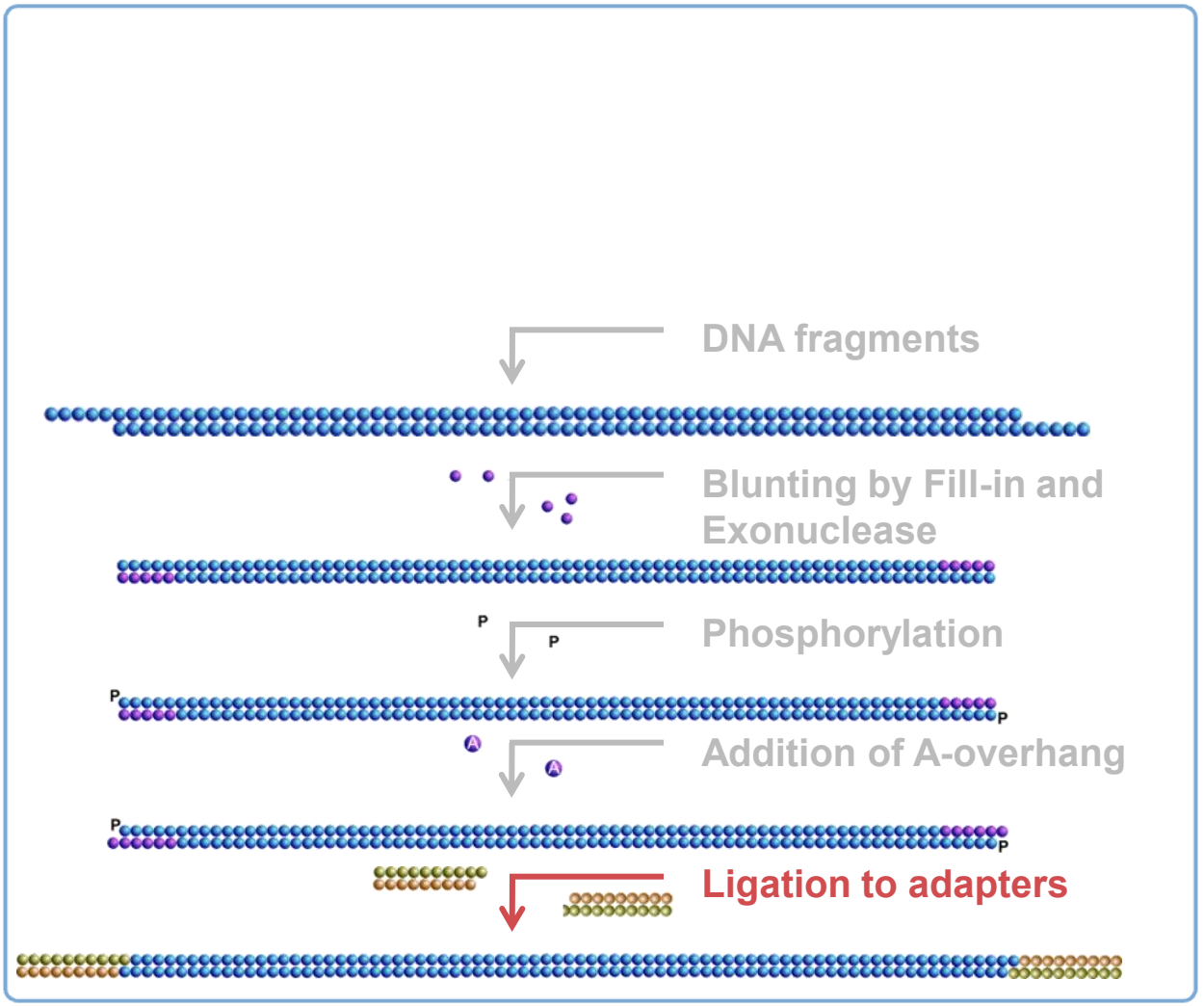
300 – 600 bp Fragments

PCR

Amplified DNA with Adapters

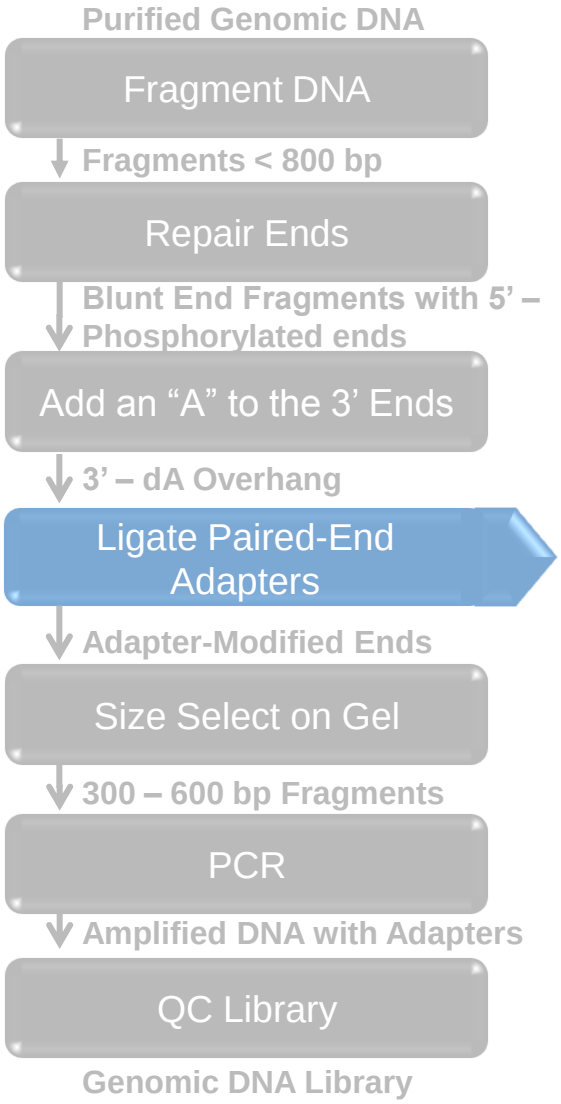
QC Library

Genomic DNA Library



BROAD INSTITUTE illumina

Step 4: Ligate Adapters



- Same library structure regardless of Application
  - gDNA
  - RNA
  - miRNA
  - ChIP-seq
  - Exon Pull-down
  - etc.

BROAD INSTITUTE illumina

Step 4: Ligate Adapters

Purified Genomic DNA

Fragment DNA

Fragments < 800 bp

Repair Ends

Blunt End Fragments with 5' – Phosphorylated ends

Add an "A" to the 3' Ends

3' – dA Overhang

Ligate Paired-End Adapters

Adapter-Modified Ends

Size Select on Gel

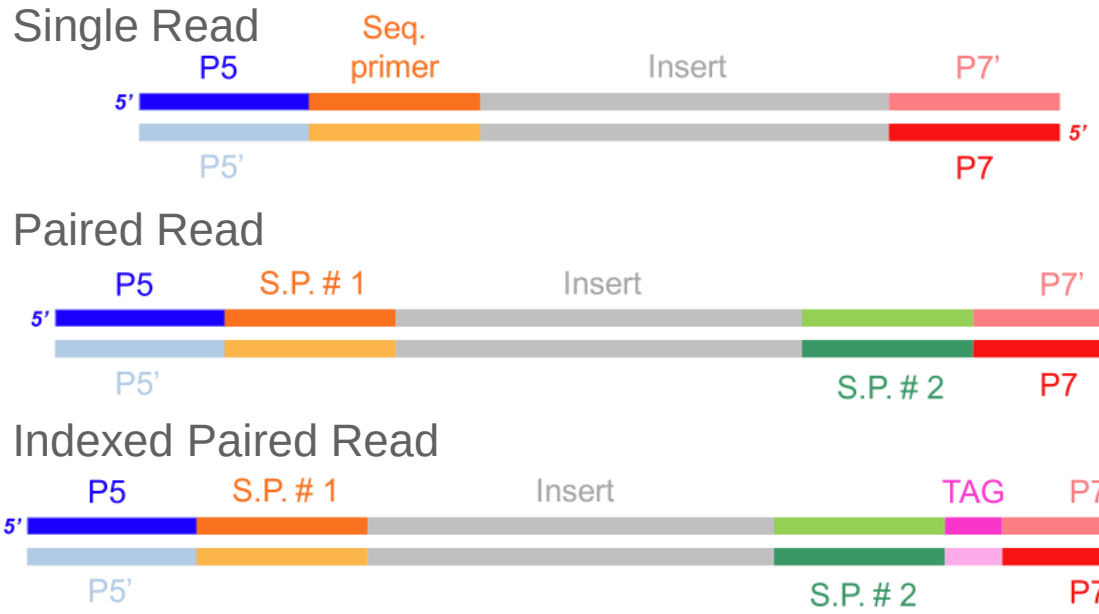
300 – 600 bp Fragments

PCR

Amplified DNA with Adapters

QC Library

Genomic DNA Library



- 10:1 molar ratio of adapter to genomic DNA insert
  - Assumes 5 µg initial input
  - Titrate for different insert amounts

BROAD INSTITUTE illumina

Choosing a Library Type

- Single read libraries:
  - Unidirectional Sequencing
  - Single Read Flowcells ONLY
  - Counting applications: ChIP or low coverage resequencing projects

- Paired end libraries:
  - Uni- OR Bi-directional (paired reads)
  - Paired End Flowcells; Single: Unidirectional only
  - Most applications, #1 whole genome shotgun assembly
  - Tailor insert size and distribution per project:
    - Tight size distribution – Assembly, structural rearrangement detection
    - Wide distribution libraries - Resequencing, high coverage

- Multiplex Paired End (aka Indexing or Barcoding)
  - Uni- OR Bi-directional
  - Allows multiple libraries per lane
  - 12 Index tags available x 8 lanes = 96 libraries per flowcell

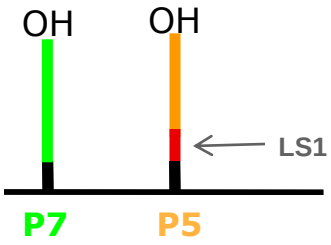
BROAD INSTITUTE illumina®

Flowcell Compatibility

Single Read Flowcell

Periodate Linearization:

- Cleaves off most of the P5 oligo



5' P-GATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG
5' ACACTCTTTCCCTACACGACGCTCTTCCGATCT

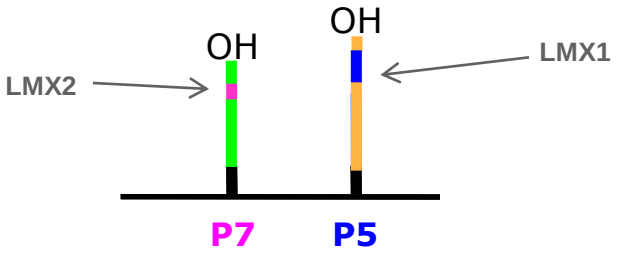
SR Adaptors

Paired End Flowcell

Linearization 1 Enzyme:

-Leaves part of P5 oligo intact for Read 2 resynthesis

Linearization 2 Enzyme



5' P-GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
5' ACACTCTTTCCCTACACGACGCTCTTCCGATCT

PE Adaptors

BROAD INSTITUTE illumina

Step 5: Gel Purify Ligation Products

Purified Genomic DNA

Fragment DNA

Fragments < 800 bp

Repair Ends

Blunt End Fragments with 5' –Phosphorylated Ends

Add an "A" to the 3' Ends

3' – dA Overhang

Ligate Paired-End Adapters

Adapter-Modified Ends

Size Select on Gel

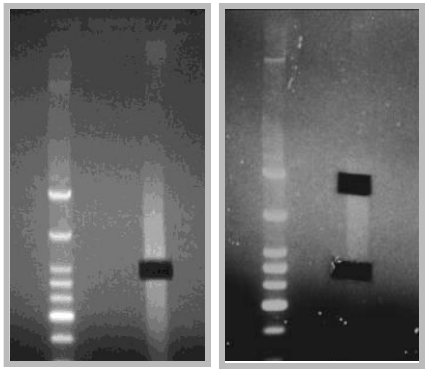
300 – 600 bp Fragments

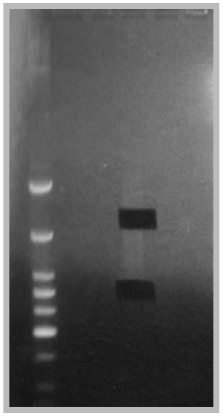
PCR

Amplified DNA with Adapters

QC Library

Genomic DNA Library

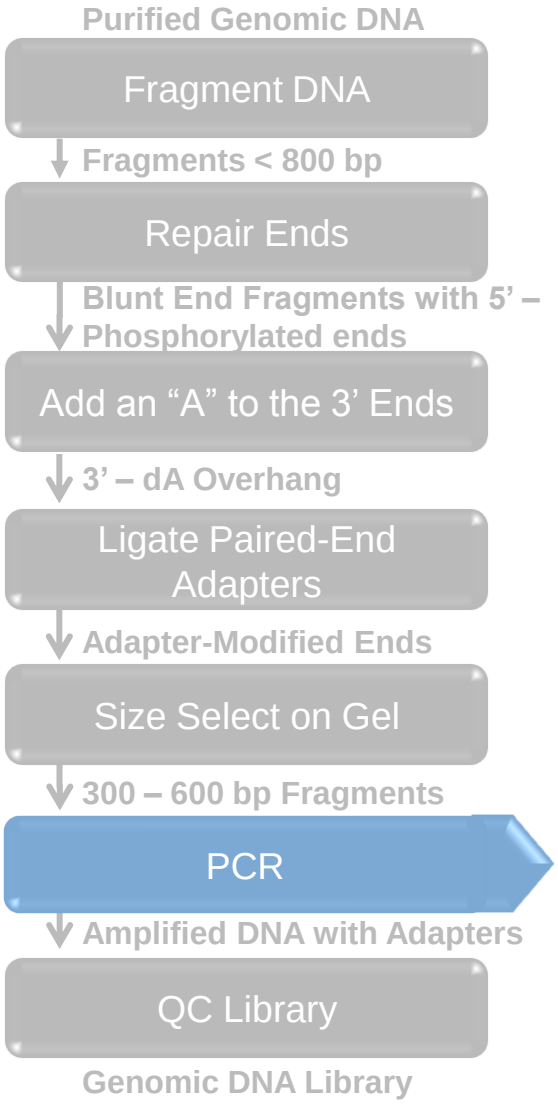




- 1µg of DNA Starting Material
  - 300/600bp area excised
- 5µg of DNA Starting Material
  - 300/600bp area excised

BROAD INSTITUTE illumina

Step 6: PCR Enrich Ligation Products



- Selectively enrich DNA fragments with adapter molecules on both ends
- Adds additional sequences to the end of the adapters for hybridization
- Amplifies the amount of DNA in the library
- Reagents:
  - Adapter Specific Primers
  - High Fidelity Enzyme (Phusion)

BROAD INSTITUTE illumina

Step 7: QC Library

Purified Genomic DNA

Fragment DNA

Fragments < 800 bp

Repair Ends

Blunt End Fragments with 5' – Phosphorylated ends

Add an "A" to the 3' Ends

3' – dA Overhang

Ligate Paired-End Adapters

Adapter-Modified Ends

Size Select on Gel

300 – 600 bp Fragments

PCR

Amplified DNA with Adapters

QC Library

Genomic DNA Library

- Quantitate by NanoDrop or Qubit
  - NanoDrop can overestimate concentration
  - Qubit or PicoGreen gives more accurate measurement
- QC by Agilent Bioanalyzer®
  - Gives size confirmation
  - Visualizes unwanted products

BROAD INSTITUTE illumina

Step 7: QC Library

Purified Genomic DNA

Fragment DNA

Fragments < 800 bp

Repair Ends

Blunt End Fragments with 5' – Phosphorylated ends

Add an "A" to the 3' Ends

3' – dA Overhang

Ligate Paired-End Adapters

Adapter-Modified Ends

Size Select on Gel

300 – 600 bp Fragments

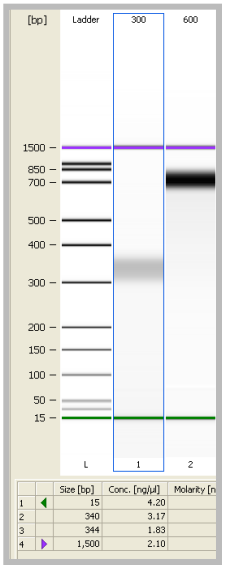
PCR

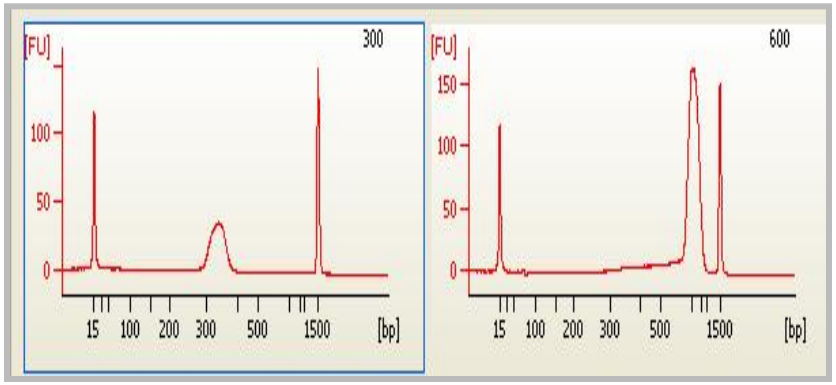
Amplified DNA with Adapters

QC Library

Genomic DNA Library

- Agilent Bioanalyzer Results
  - Lane 1: 300 bp library
  - Lane 2: 600 bp library





BROAD INSTITUTE illumina

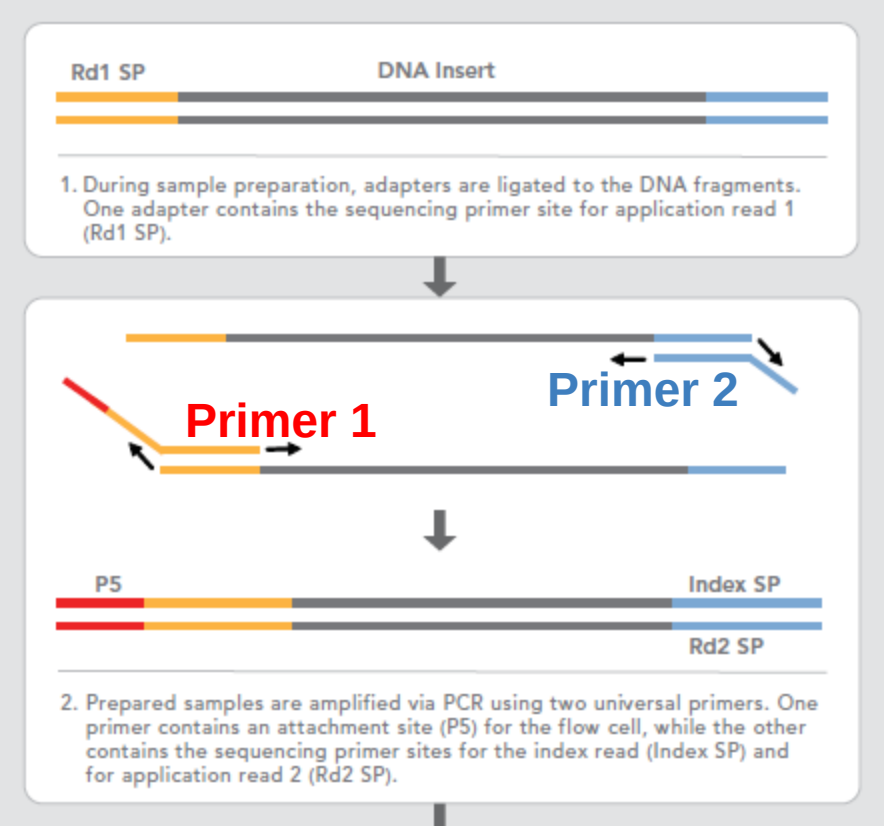
Other DNA Applications

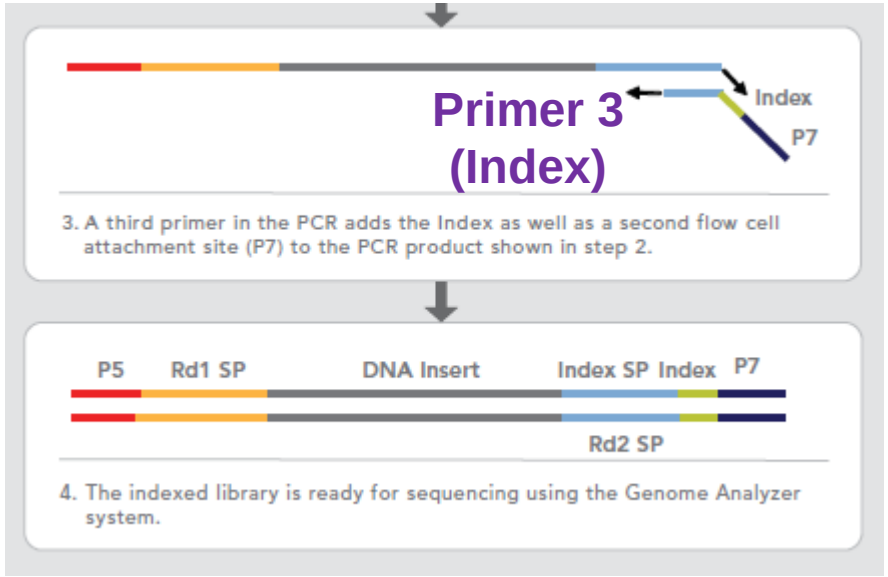
- Multiplexing (aka Indexing/Barcoding)
- Large Insert Mate Pair

BROAD INSTITUTE illumina®

Multiplex Paired End Libraries

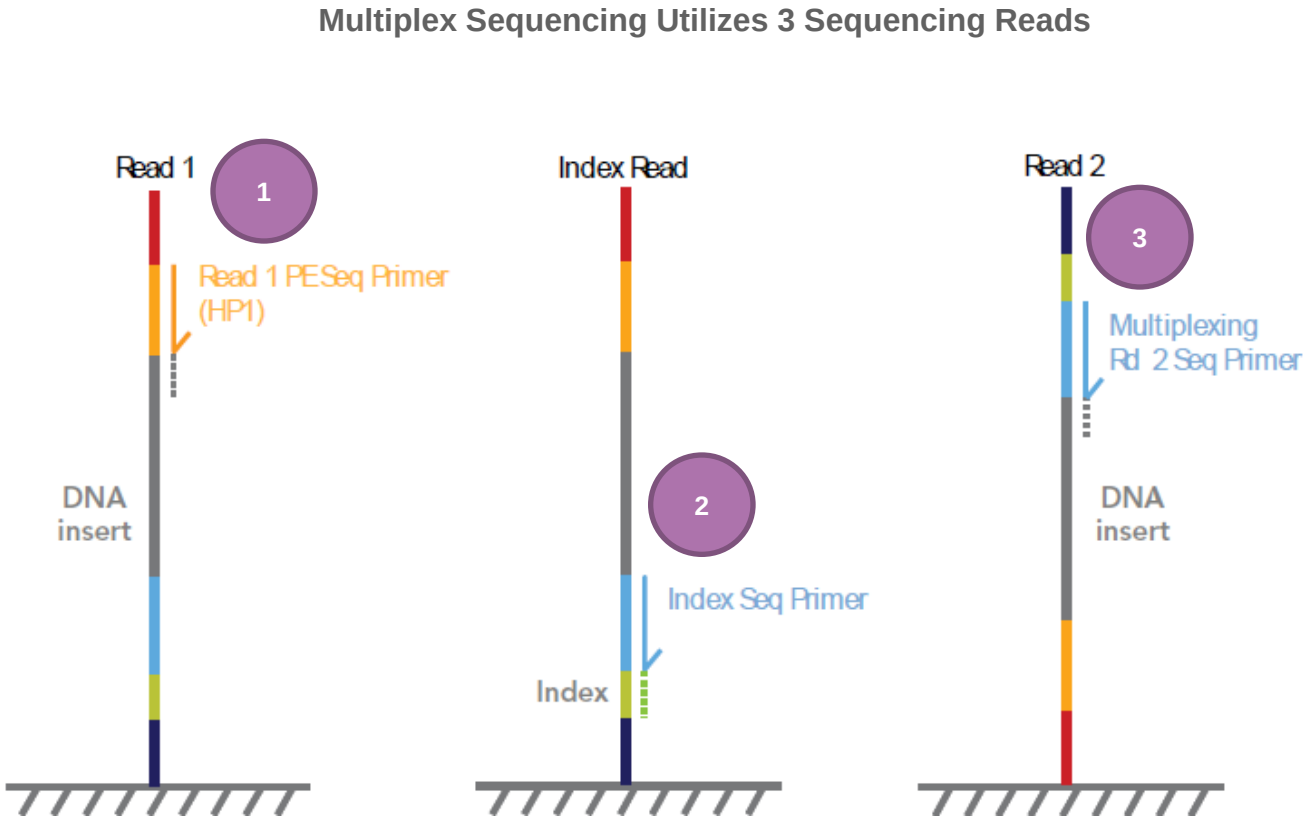
Index Tag is incorporated during Sample Prep PCR, using a 3-primer system





BROAD INSTITUTE illumina

Multiplex Paired End Libraries



Long Insert Mate Pair

- Allows gap sizes of 2-5kb
- Useful for whole genome shotgun assembly; connecting contigs
- Detection of structural/chromosomal rearrangements

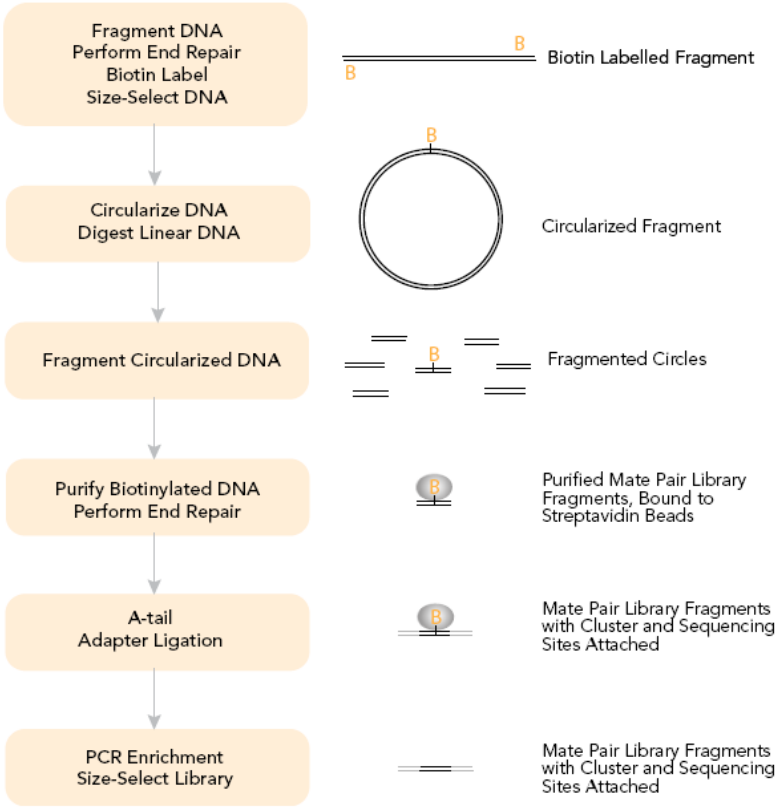


Figure 1 *Mate Pair Library Preparation Overview*

BROAD INSTITUTE illumina

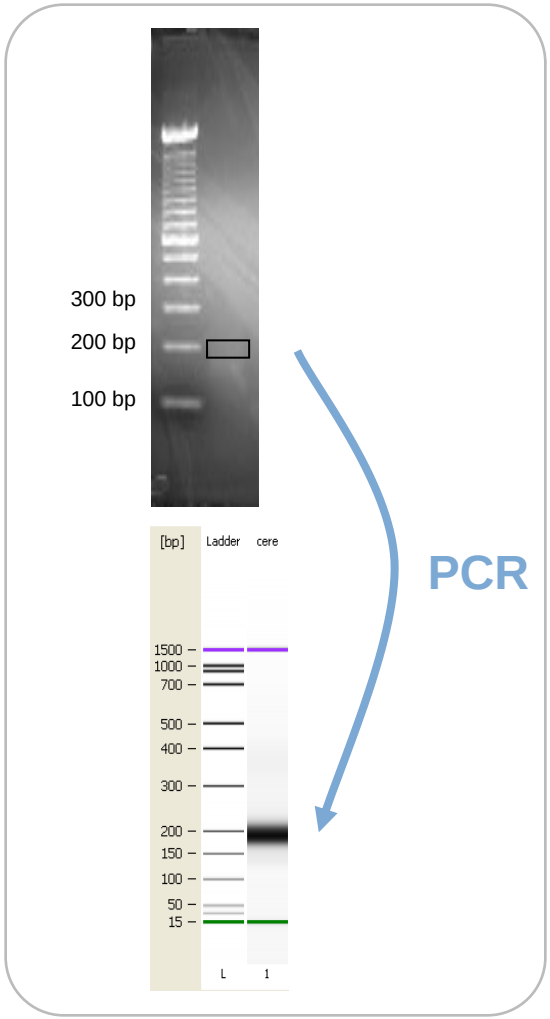
RNA Applications Sample Prep

- mRNA-Seq
- Small RNA Seq

BROAD INSTITUTE illumina

mRNA-Seq Sample Prep

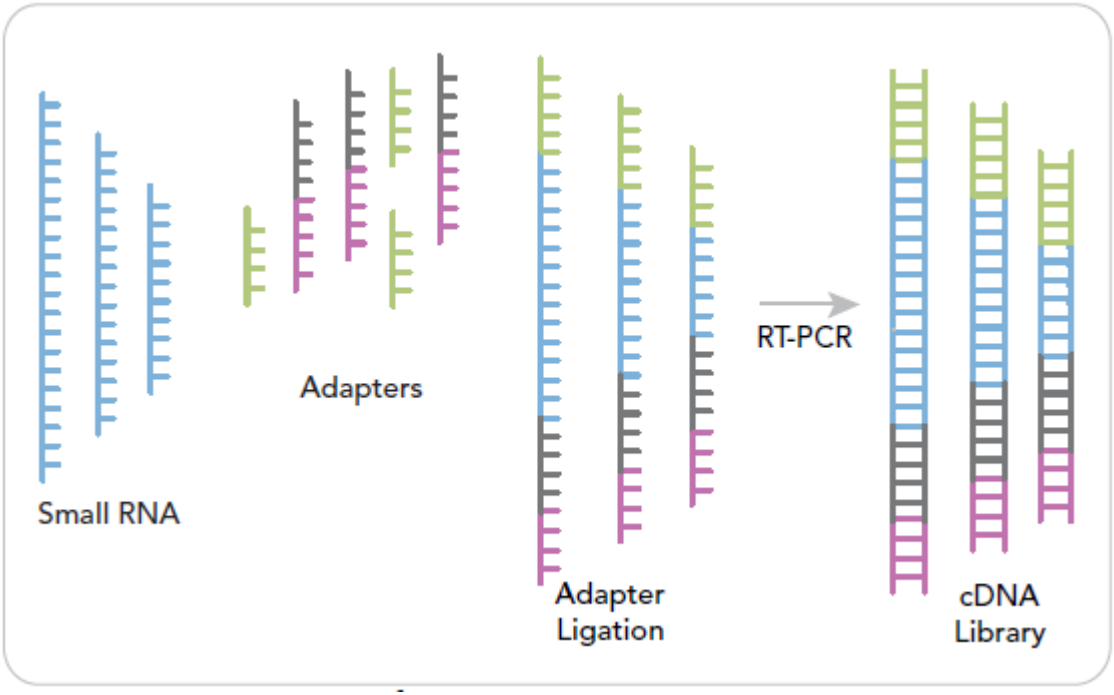
- Isolate poly-A containing mRNA
  - Magnetic oligo dT beads
- Randomly fragment RNA
- Random prime mRNA → cDNA
- Make 2nd strand cDNA
- Ligate sequencing adapters
- Size select on gel
- Enrich with 15 cycles of PCR



BROAD INSTITUTE illumina

Diagram of Ligation Reaction

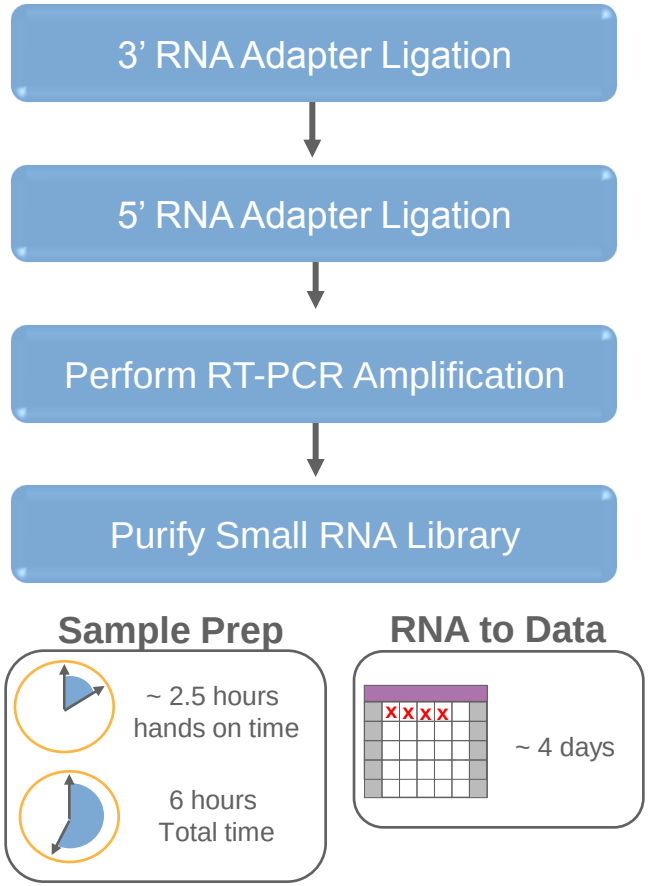
- Total RNA
- 3’ adapter ligation
- 5’ adapter ligation
- RT-PCR
- Sequencing



BROAD INSTITUTE illumina

Small RNA V1.5 Sample Prep

- 6 hour library preparation
  - Shortest time from library to sequence
  - 2.5 hrs hands on time (6 hr total)
  - Single gel excision
- NO need for small RNA purification!
  - Starts directly from total RNA
- Reduced RNA input requirements
  - 1.0 µg total RNA input
    - Titrated to 200 ng total
- Strand specific information
  - Discover and profile microRNA
  - Sequence other non-coding RNAs
  - Can be used with stranded mRNA-Seq



BROAD INSTITUTE illumina

Sample Preparation Best Practices

Sample Preparation

- Be familiar with the protocol before starting it
  - Have all user-supplied reagents and equipment
- Use high-quality starting material
  - QC and quantify material before starting protocol
  - Aim for middle to high end of starting material requirements if possible
  - Very important for RNA applications
- Use proper material handling techniques
  - Prevent contamination and degradation
  - Particularly important for RNA applications

BROAD INSTITUTE illumina

Sample Preparation Best Practices

Sample Preparation

- Use positive controls if necessary for troubleshooting
- Quantitate and QC final library
  - This is essential for cluster lawn quality and sequencing success

Sample Preparation

Best Practices from The Broad Institute

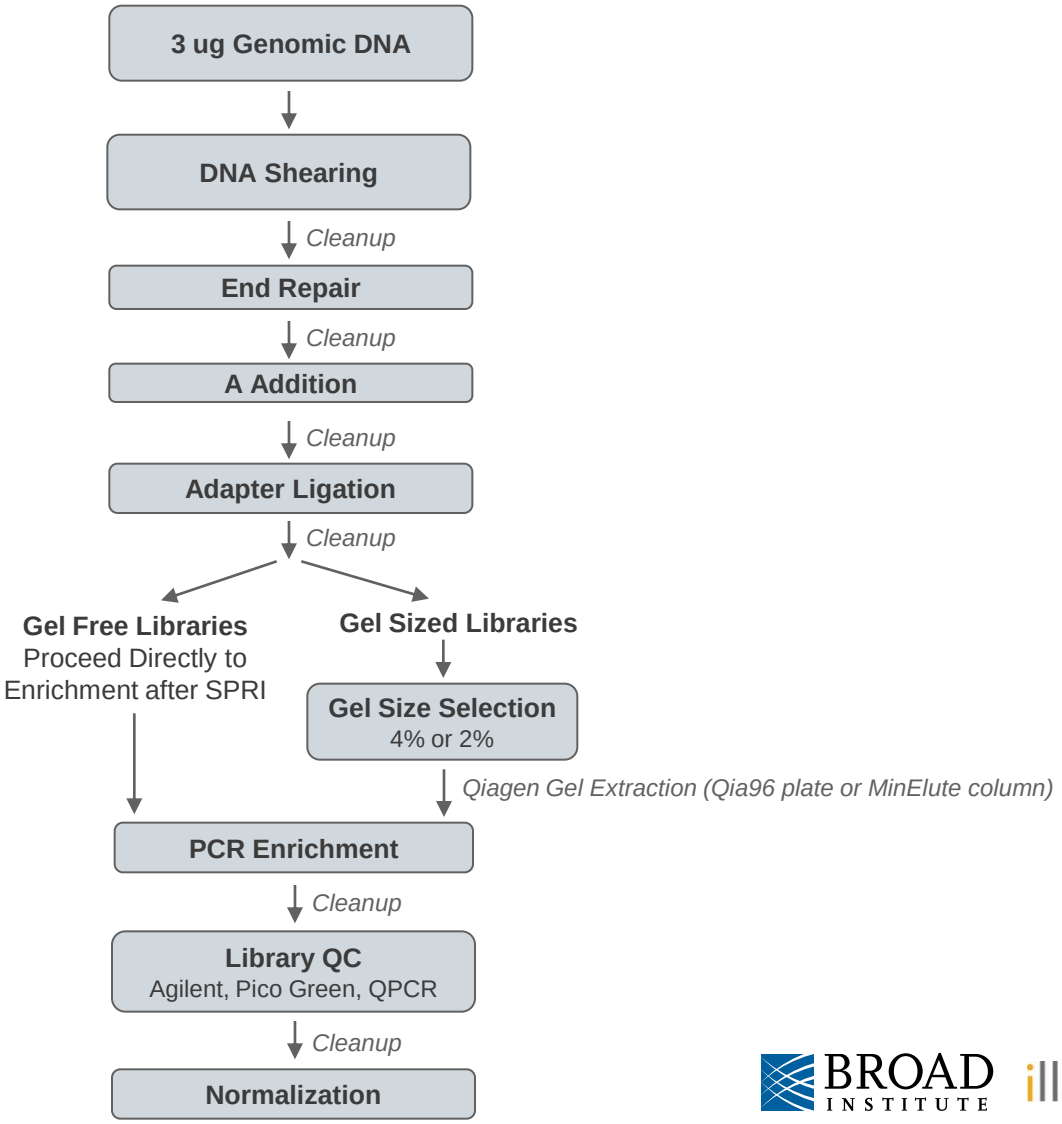
BROAD INSTITUTE illumina

Module Objectives – Sample Preparation

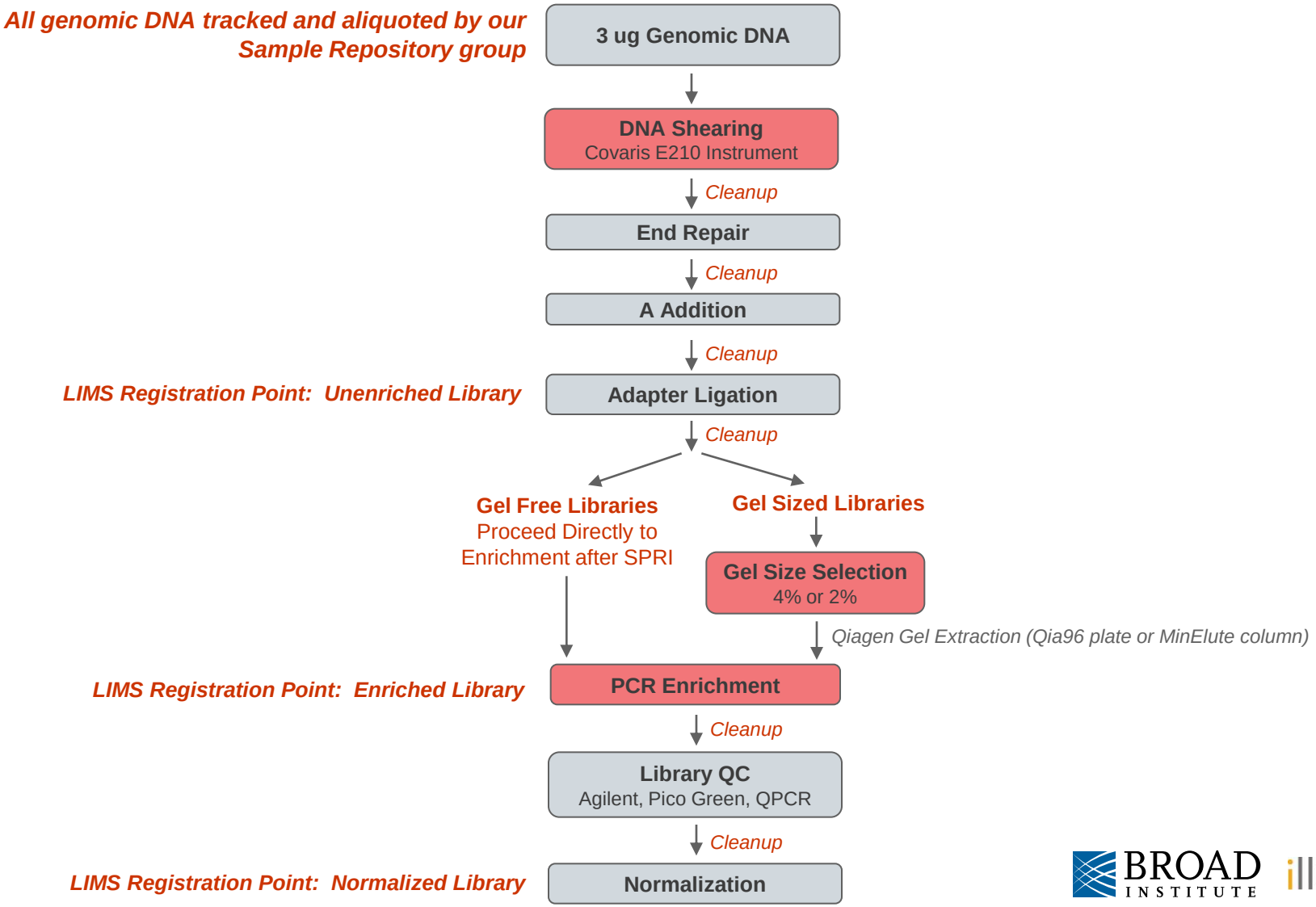
- **Overview of Broad's sample preparation workflow**
- **How we track samples through the lab**
- **Broad developed modifications to sample preparation and their application in our protocols:**
  - Shearing methods
  - Enzymatic reaction clean up methods
  - Size selection methods
    - Both agarose gel and gel-free methods
  - PCR enrichment protocols
- **Quality control methodology for sample preparation**
  - Overview of sequencing metrics that can be used to assess sample preparation quality

BROAD INSTITUTE illumina

Sample Prep Workflow at the Broad

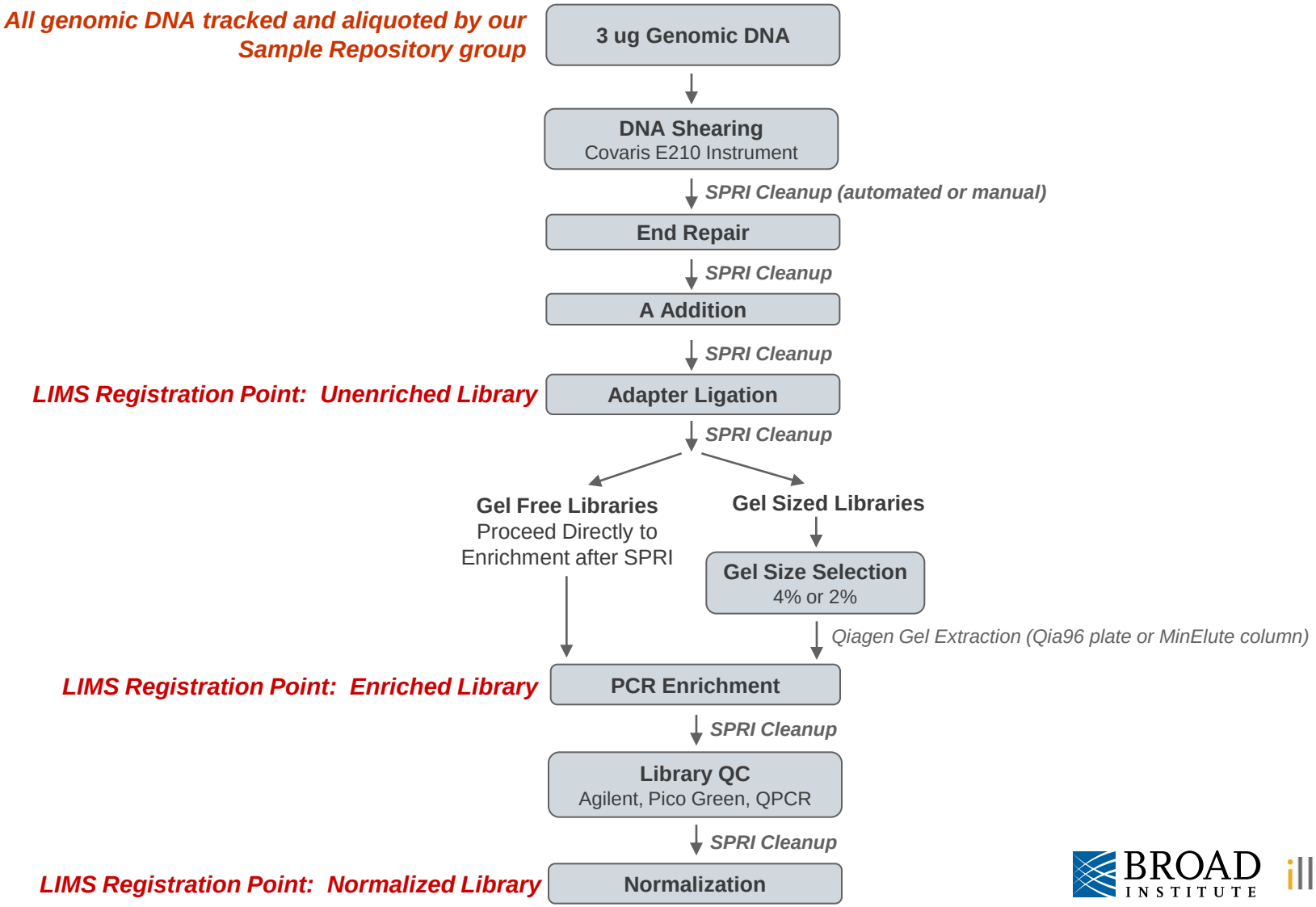


Sample Prep Workflow at the Broad



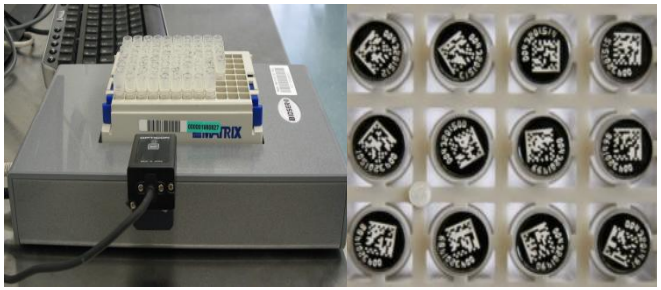
Sample Tracking

Sample Tracking

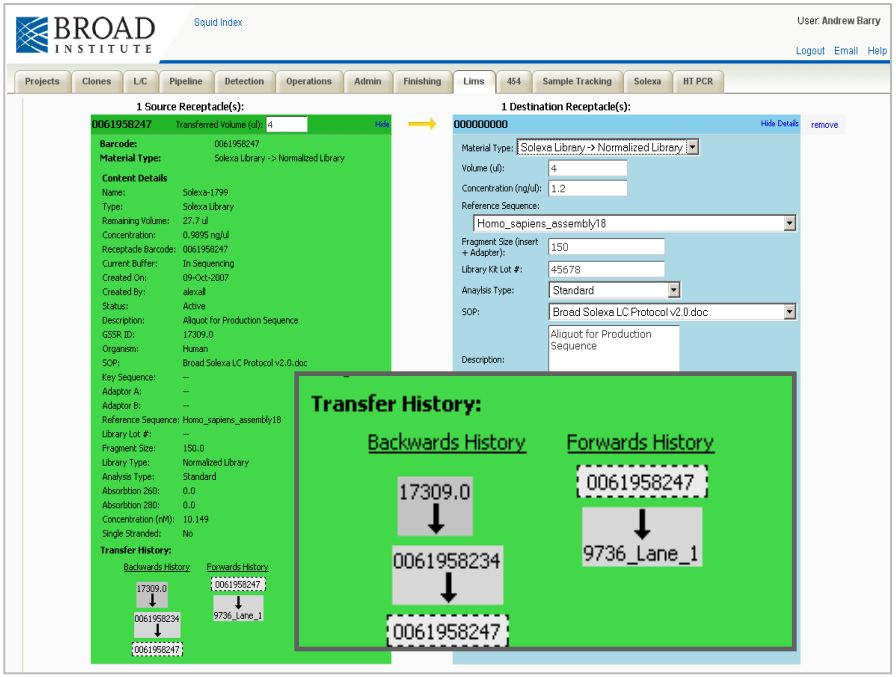


Sample Tracking in a High Throughput World

- DNA stored in 2D barcoded tubes
  - Source Genomic DNA
  - Unenriched libraries
  - Enriched libraries
  - Normalized libraries



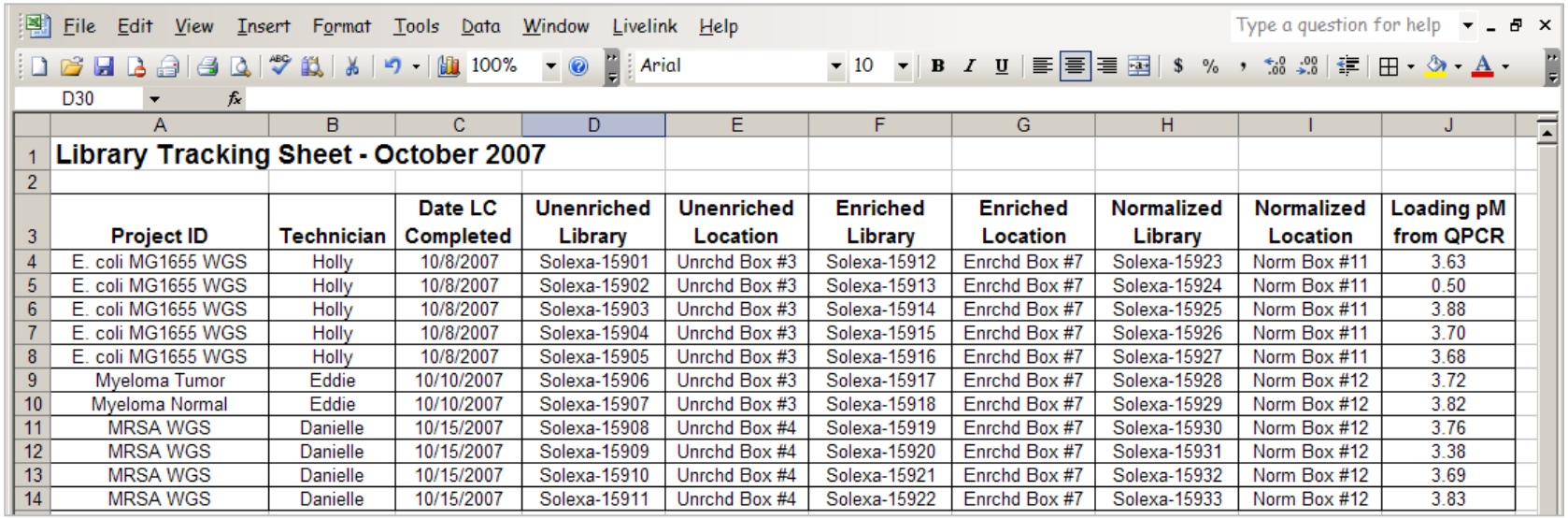
- All transfers and details are tracked via web-based LIMS



- Tube picking system error-proofs selecting libraries

BROAD INSTITUTE illumina®

Sample Tracking at Lower Throughputs



File Edit View Insert Format Tools Data Window Livelink Help — Type a question for help

D30

Library Tracking Sheet - October 2007

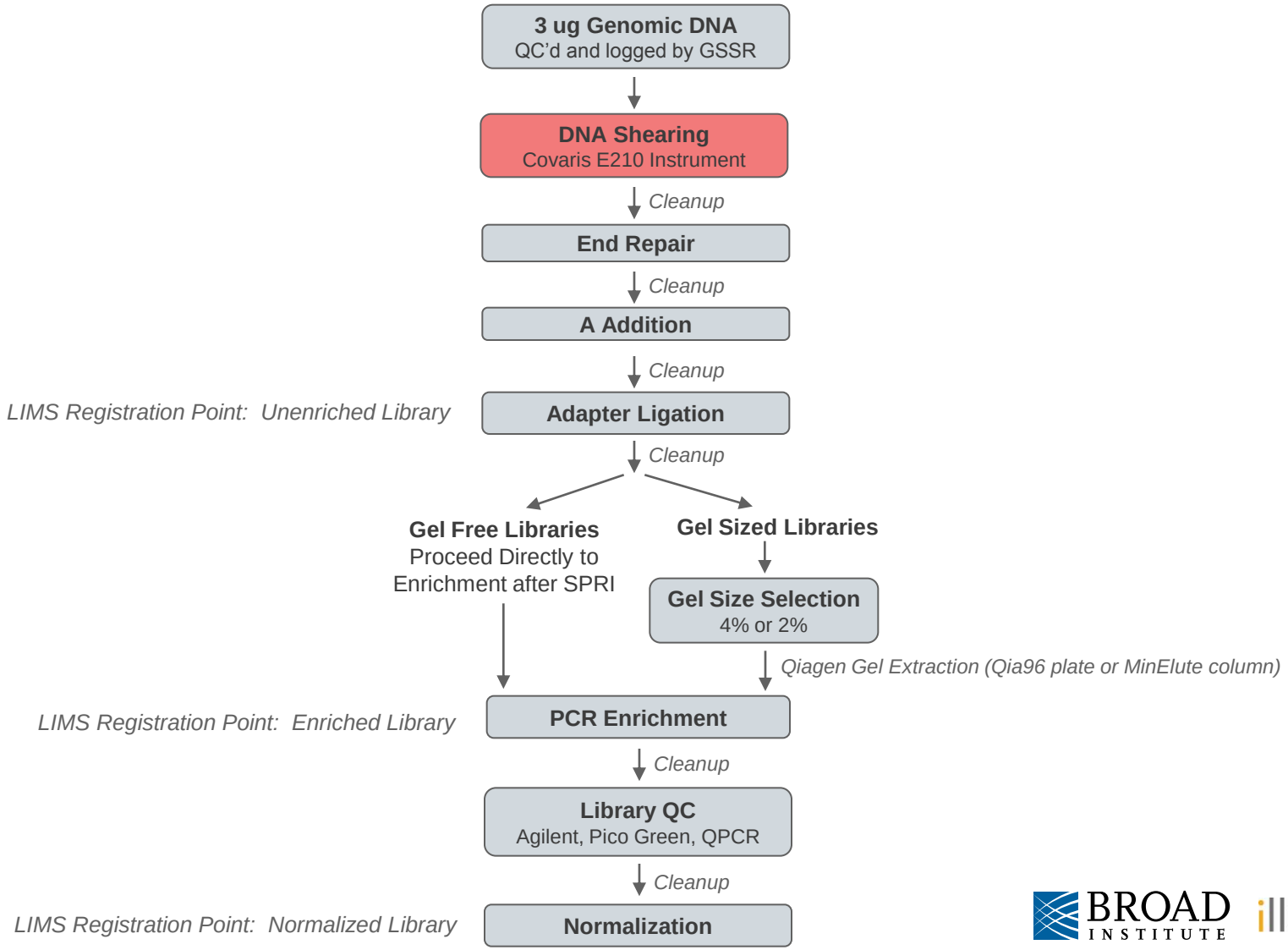
Project ID	Technician	Date LC Completed	Unenriched Library	Unenriched Location	Enriched Library	Enriched Location	Normalized Library	Normalized Location	Loading pM from QPCR
E. coli MG1655 WGS	Holly	10/8/2007	Solexa-15901	Unrchd Box #3	Solexa-15912	Enrchd Box #7	Solexa-15923	Norm Box #11	3.63
E. coli MG1655 WGS	Holly	10/8/2007	Solexa-15902	Unrchd Box #3	Solexa-15913	Enrchd Box #7	Solexa-15924	Norm Box #11	0.50
E. coli MG1655 WGS	Holly	10/8/2007	Solexa-15903	Unrchd Box #3	Solexa-15914	Enrchd Box #7	Solexa-15925	Norm Box #11	3.88
E. coli MG1655 WGS	Holly	10/8/2007	Solexa-15904	Unrchd Box #3	Solexa-15915	Enrchd Box #7	Solexa-15926	Norm Box #11	3.70
E. coli MG1655 WGS	Holly	10/8/2007	Solexa-15905	Unrchd Box #3	Solexa-15916	Enrchd Box #7	Solexa-15927	Norm Box #11	3.68
Myeloma Tumor	Eddie	10/10/2007	Solexa-15906	Unrchd Box #3	Solexa-15917	Enrchd Box #7	Solexa-15928	Norm Box #12	3.72
Myeloma Normal	Eddie	10/10/2007	Solexa-15907	Unrchd Box #3	Solexa-15918	Enrchd Box #7	Solexa-15929	Norm Box #12	3.82
MRSA WGS	Danielle	10/15/2007	Solexa-15908	Unrchd Box #4	Solexa-15919	Enrchd Box #7	Solexa-15930	Norm Box #12	3.76
MRSA WGS	Danielle	10/15/2007	Solexa-15909	Unrchd Box #4	Solexa-15920	Enrchd Box #7	Solexa-15931	Norm Box #12	3.38
MRSA WGS	Danielle	10/15/2007	Solexa-15910	Unrchd Box #4	Solexa-15921	Enrchd Box #7	Solexa-15932	Norm Box #12	3.69
MRSA WGS	Danielle	10/15/2007	Solexa-15911	Unrchd Box #4	Solexa-15922	Enrchd Box #7	Solexa-15933	Norm Box #12	3.83

- Implementing a library numbering system allows for easy tracking, even without barcodes
- Keeping Excel tracking sheets can associate libraries to each other and to project, assisting in downstream troubleshooting
  - Records who made the libraries, prep date, lot #s, etc.
  - Identifies where in the freezer the tubes are stored

BROAD INSTITUTE illumina

Shearing of Genomic DNA

Shearing of Genomic DNA



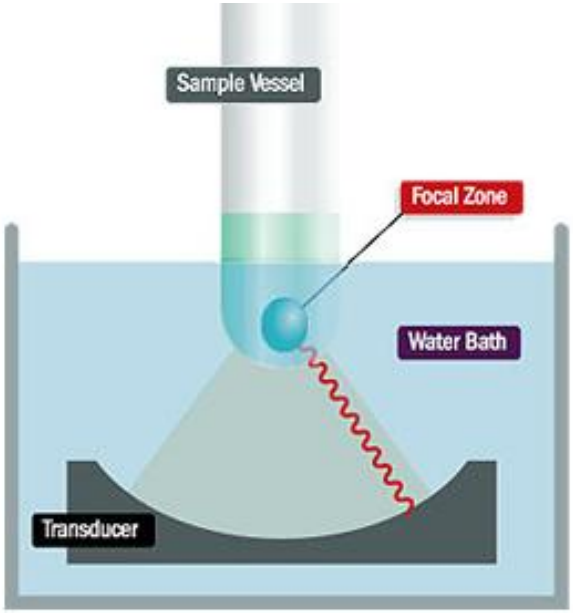
DNA Shearing with Covaris

- Founded in 1999, Covaris, Inc. developed a line of instruments capable of providing controlled delivery of acoustic energy to closed vessels.
- Originally, Covaris targeted the chemical industry in order to facilitate mixing, homogenizing, or pulverization of chemical compounds.
- We were the first to work with Covaris to develop methods to apply their “adaptive focused acoustics” technology to the process of shearing DNA for next generation sequencing applications.
- Main reason for moving to Covaris was the throughput:
  - Up to 96 samples can be loaded into the machine at a time in the E210 plate based model.

BROAD INSTITUTE illumina

DNA Shearing with Covaris

- How it works:
  - Acoustic ultrasonic waves aimed at glass shearing vessels containing DNA cause “cavitation” within the sample.
  - This extremely rapid creation and high-energy collapsing of tiny bubbles leads to the mechanical shearing of DNA.
  - We will see a demo in the Lab Section.



E210 Plate Based Model

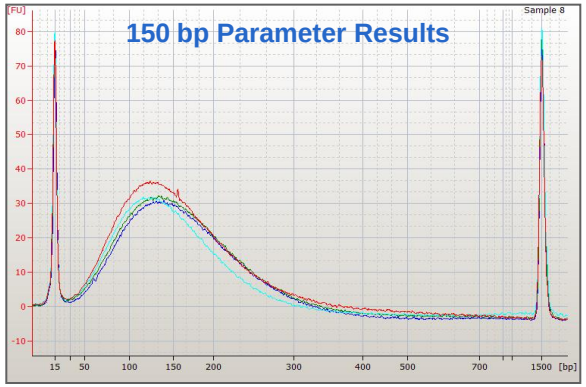
S2 Single Tube Based Model

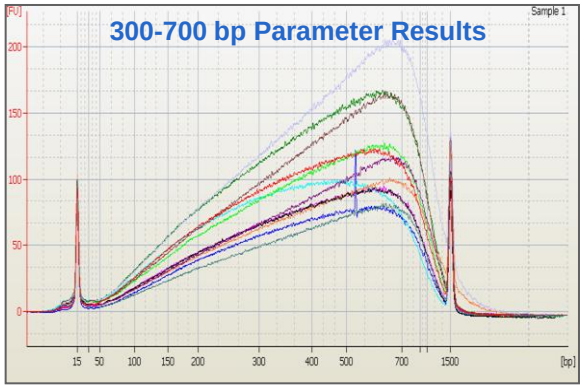
BROAD INSTITUTE illumina

DNA Shearing with Covaris

- We have performed a multitude of experiments to understand how the parameters can be manipulated to produce DNA in various sizes:

Desired Fragment Size	Volume	Time (T)	Duty Cycle	Intensity	Cycles Per Burst (CPB)	Z Axis (mm)
150bp	100	300 sec	20%	5	200	0mm
200bp	100	90 sec	20%	5	200	0mm
200bp	50	360 sec	5%	5	200	0mm
300-700bp	100	30 sec	20%	5	200	0mm
300-700bp	50	300 sec	1%	5	200	0mm





- *You will see more examples of these data in the lab section.*

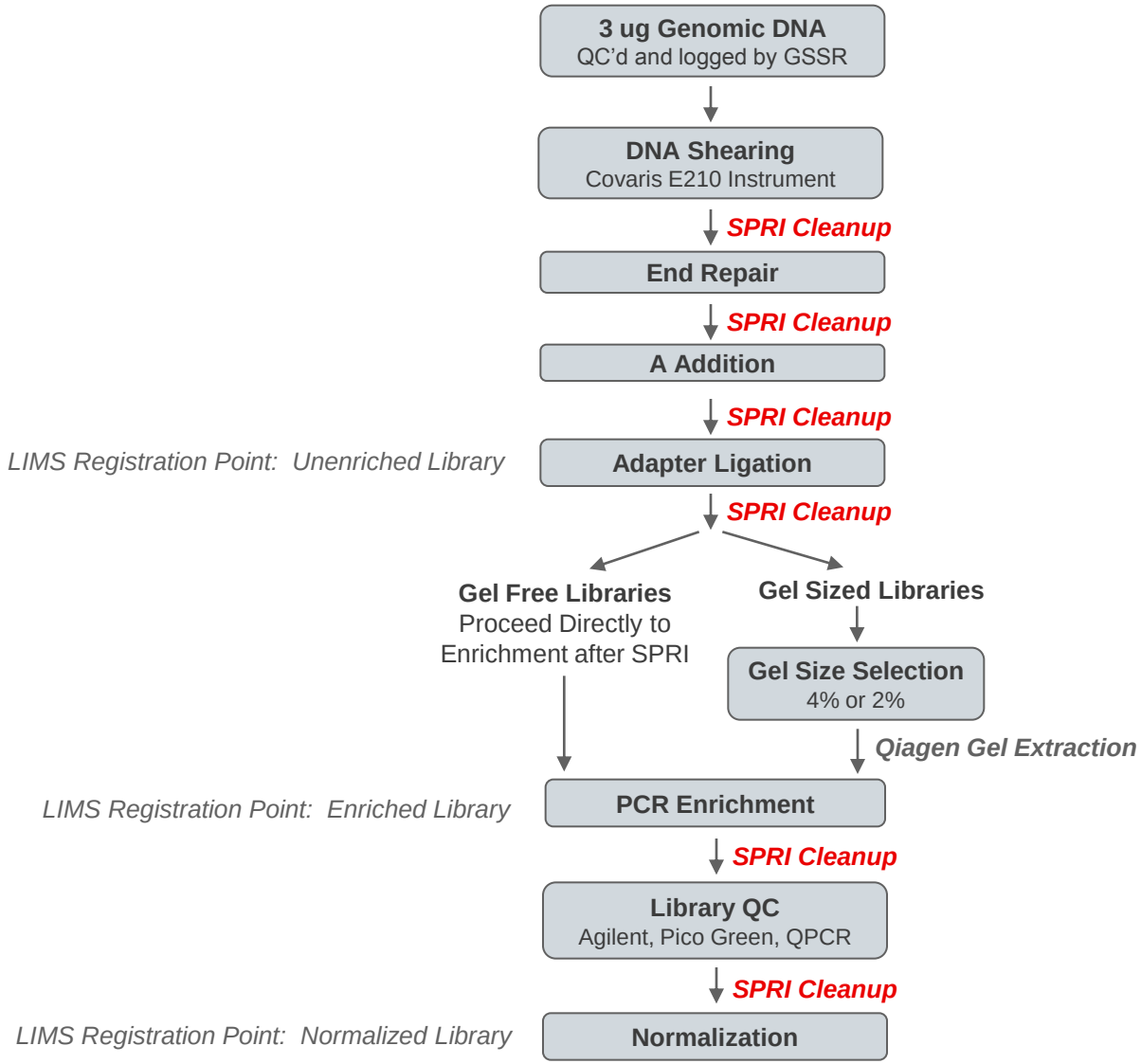
Choosing a DNA Shearing Method

	DNA in Desired Range	Reproducible	Customizable Settings	No Additional Materials Cost	No Additional Instrument Required	Multiple Samples at a Time	Compatible with automation
Nebulizer	✓	✓	*Limited - can vary time and pressure*	✓	✓		
Covaris S2 Single Tube	✓	✓	✓				
Covaris E210 Plate Based	✓	✓	✓			✓	✓

BROAD INSTITUTE illumina®

Reaction Cleanups

Reaction Cleanups



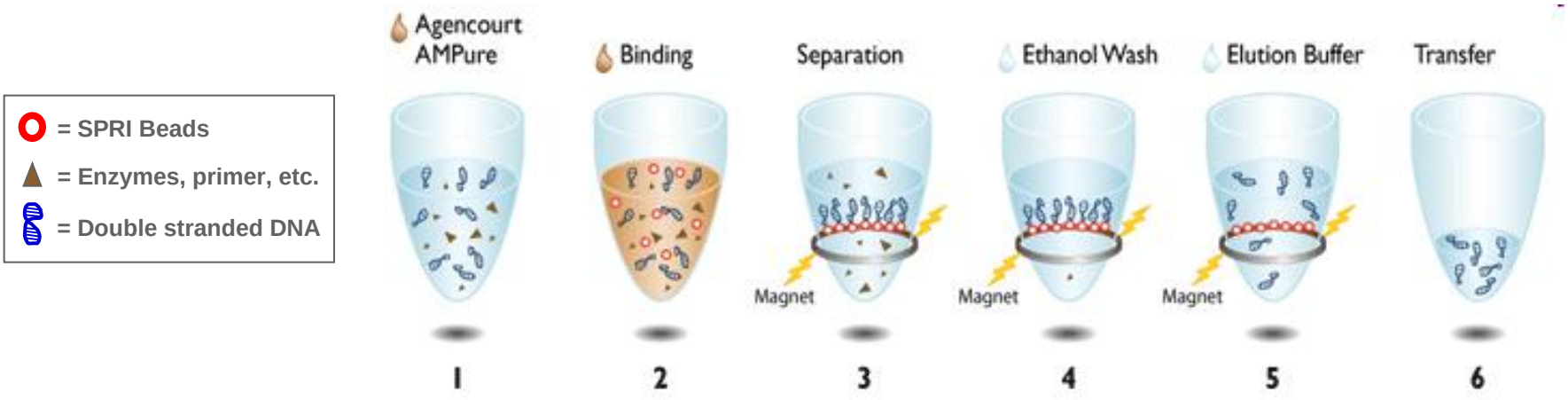
Reaction Cleanup Methods

- Column based cleanups with Qiagen MinElute columns
  - Reliable and reproducible, ~80% recovery
  - Easy to perform
  - Manual - ideal for lower throughput
    - Not compatible with automation: reasonable set size <12
    - Tubes more prone to sample swapping than samples in a plate
- Magnetic bead based cleanup methods
  - Reliable and reproducible, 80-90% recovery
  - Easy to perform
  - Can be done manual or automated
    - Manual: small sets of 8-12 can be reasonably done by hand in tubes
    - Automated: up to 96 can be done at a time using a liquid handler
- **Broad has moved away from column cleanups to magnetic bead based cleanups to accommodate higher throughput**

BROAD INSTITUTE illumina

SPRI: An Alternative to Column Cleanups

- SPRI: **S**olid **P**hase **R**eversible **I**mmobilization (*AMPure by Agencourt*)
  - SPRI technology is a magnetic bead based DNA clean up method
  - Beads are specific for double stranded DNA
  - Enzymes, excess primer, etc. are washed away & DNA is then eluted
  - Reference: Hawkins TL et al. Nucleic Acids Res. 1994 Oct 25;22(21):4543-4.
- How the SPRI chemistry works:

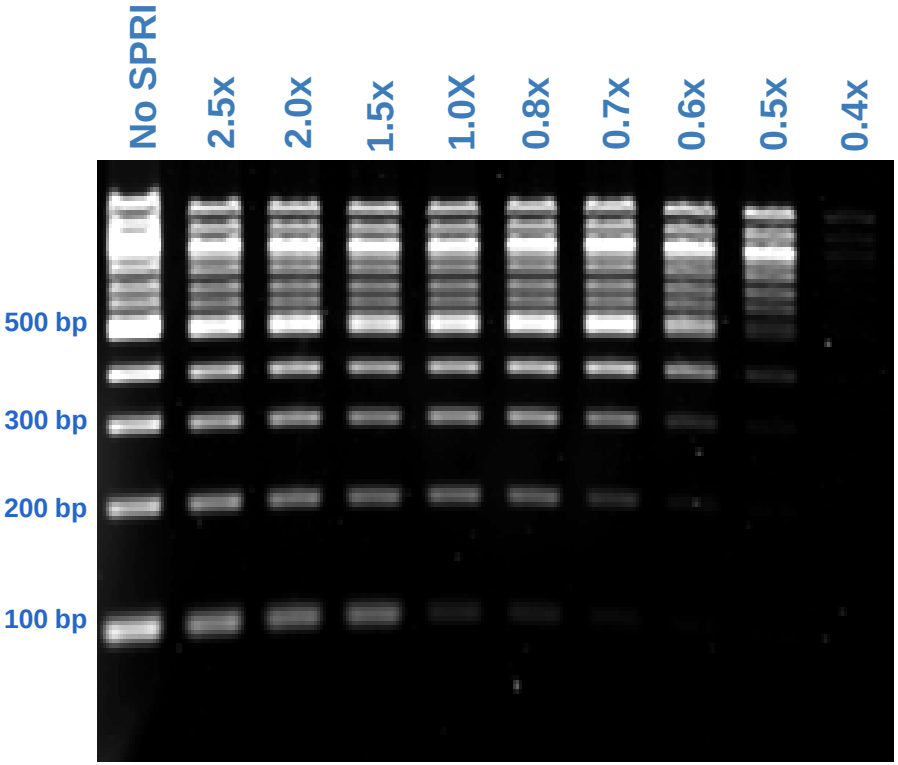


Courtesy of Agencourt

BROAD INSTITUTE illumina

SPRI: An Alternative to Column Cleanups

- The length of fragments that SPRI binds can be controlled by varying the amount that is added to the reaction
- The “X” of SPRI refers to the volume to volume ratio of SPRI to DNA
  - 1X SPRI is a 1:1 vol of SPRI:DNA
- The **lower** the X of SPRI, the **higher** the size cut off becomes
  - See titration at right
  - 0.7X SPRI removes fragments below 150 bp such as adapter dimer
- As we will see in the size selection section, this property can be used to perform gel free library sizing



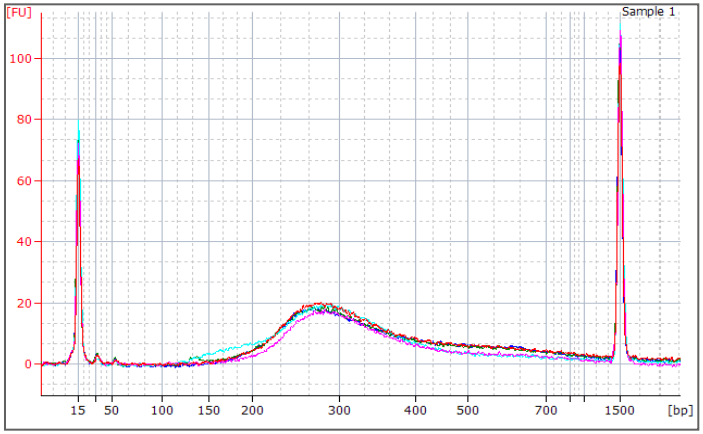
BROAD INSTITUTE illumina

Application of SPRI - High Throughput Libraries

- To fill our sequencing capacity, we needed to be able to do the following:
  - Make 96 libraries or more per week
  - Start with 3 ug of DNA
  - Average final fragment size ~300 bp but tight sizing NOT required

- Implementing Covaris shearing and SPRI cleanups made this possible:
  - Covaris shear of 96 samples
  - Automated SPRI cleanups on the Velocity 11 Bravo liquid handler
  - No size selection method
    - 150 bp Covaris shearing
    - 0.7X SPRI to remove dimer after adapter ligation

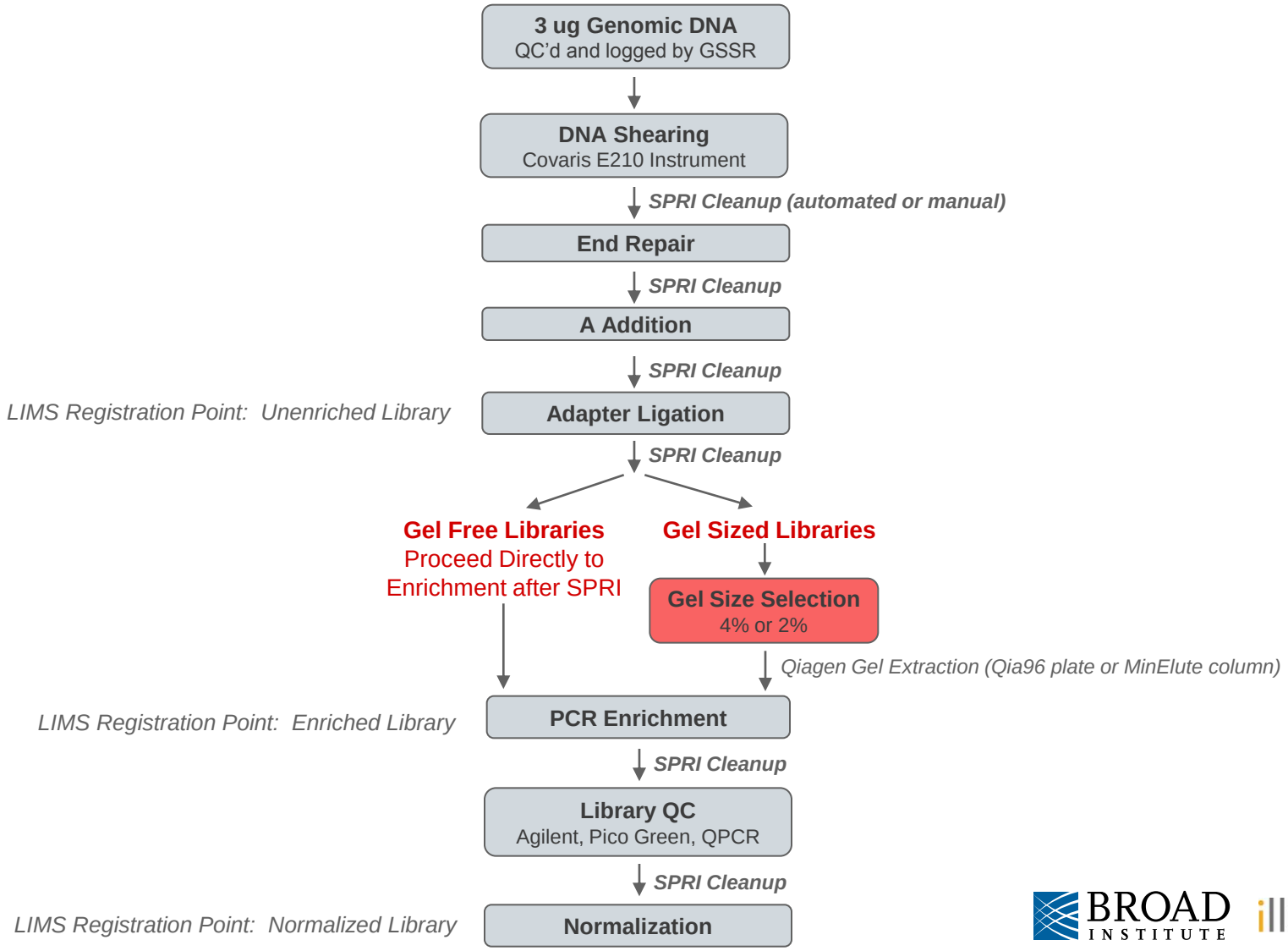
Completed Library Agilent QC



BROAD INSTITUTE illumina

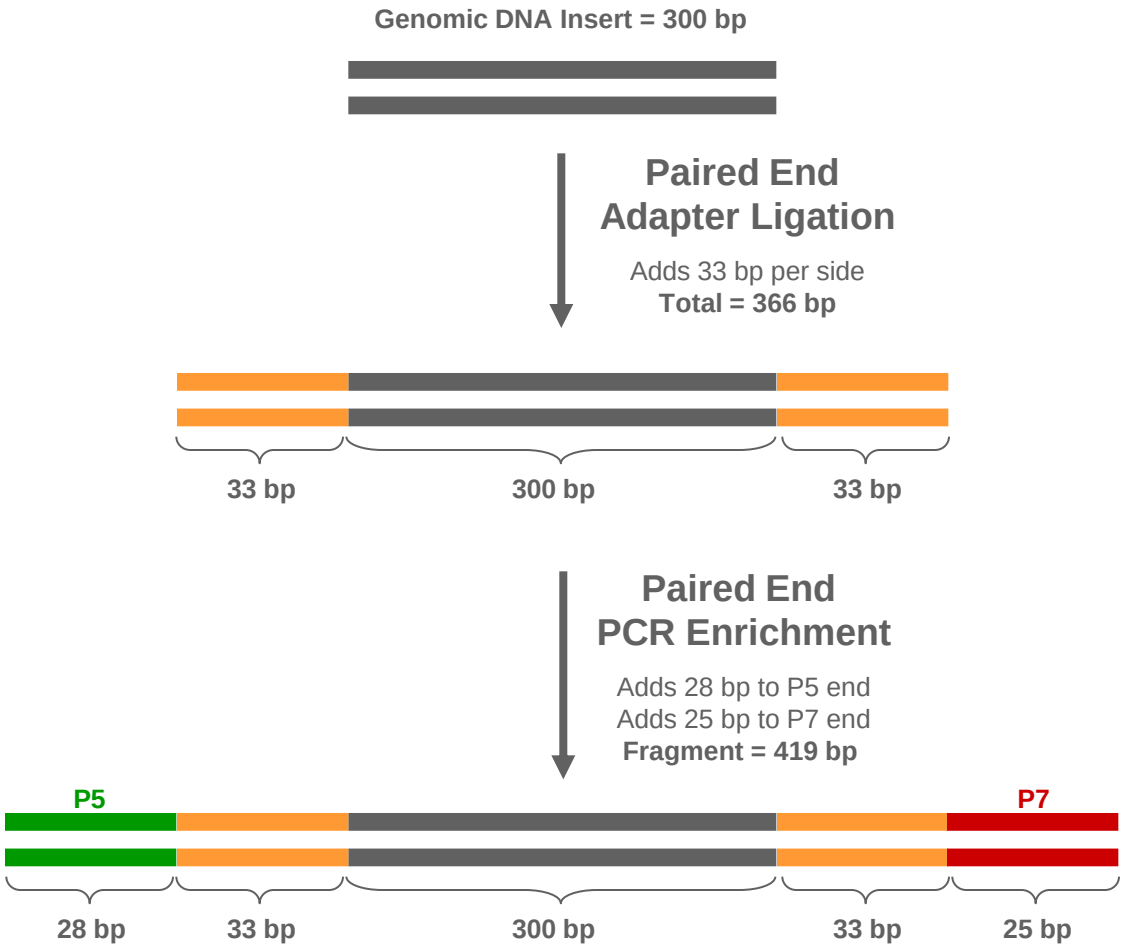
Size Selection Methods

Size Selection



BROAD INSTITUTE illumina

Insert Size vs. Library Fragment Size



Choosing a Library Fragment Size

- A library fragment size of ~400 bp is ideal for optimal cluster density as well as for sequencer optics and cluster-finding algorithms
- However, the sequencer can tolerate fragment sizes up to 800 bp
  - Fragment size can be tailored to suit the needs of a given project
- Desired read length determines the minimum insert size
  - For paired end, the targeted insert size should usually be at least twice the read length to avoid reading into the adapter or sequencing the same stretch
    - For 36 base paired sequencing: at least 72 bases
    - For 76 base paired sequencing: at least 152 bases
- **Larger fragments → Larger clusters → Lower max density**
- **Wider range → Wider variety of cluster sizes → Lower max density**

BROAD INSTITUTE illumina

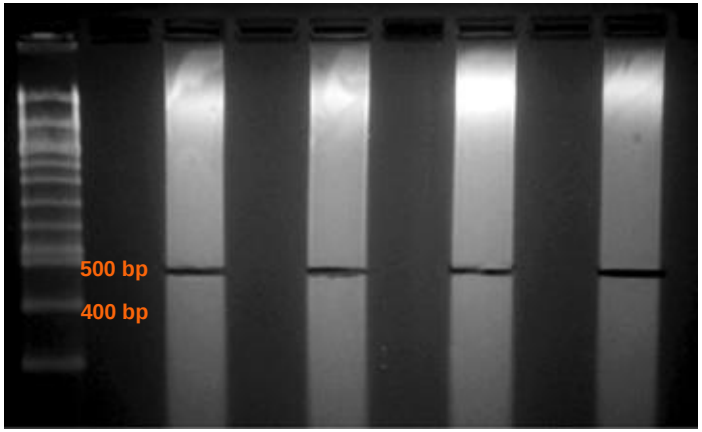
Choosing a Size Selection Method

- Option 1: Agarose gels
- Option 2: Double-sided SPRI
- Option 3: No size selection

Option 1: Agarose Gels

- Gel size selection is necessary to achieve size deviations of < 30%.
- Broad's recommendations for gel conditions:
  - **For fragments <400 bp: 4% NuSieve 3:1 TAE gel, 85 volts, 3 hours, room temp**
  - **For fragments >400 bp: 2% SeaPlaque TAE gel, 110 volts, 3 hours, 4°C**
- By taking ~2 mm gel slices at desired size range, a size distribution of +/- 10% or less can be achieved:

We will see a lab demo of a gel cutting tool we have developed to make these tight cuts easier.

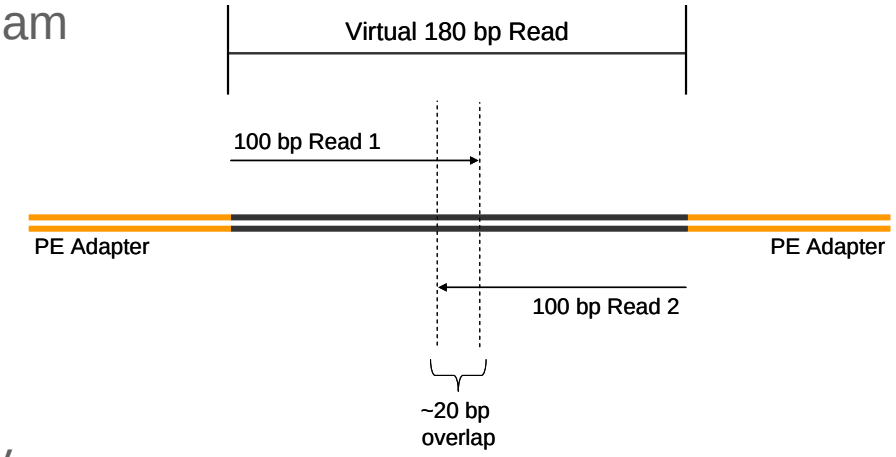


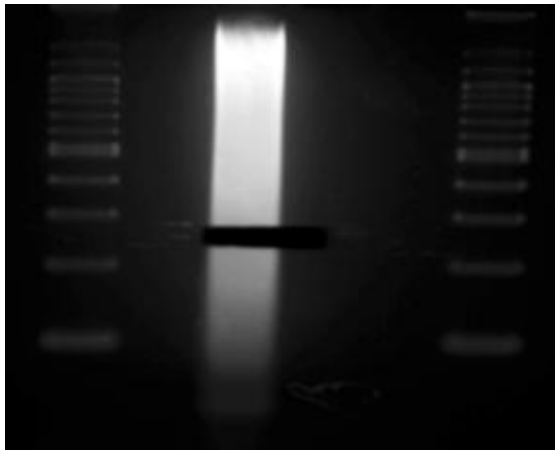
Example of a tight cut library targeting fragment size of 490 bp.

BROAD INSTITUTE illumina

Application of Tight Cuts - 180 bp Libraries

- We were requested by our Assembly team to make libraries with inserts that were EXACTLY 180 bp with less than a 10% deviation
  - 2x101 run = virtual 180 bp read
  - 180 bp virtual read more robust to assemble than 100 bp alone
- To accomplish this, we take 3 extremely tight gel slices at 246 bp after adapter ligation (180 bp insert + 66 bp adapters)
  - Use a gel cutting tool that holds 4 razor blades - Will see a demo in lab section
  - Perform “high diversity” PCR protocol
  - QC size & distribution by Agilent chip
    - After PCR, should be 299 bp
- Select library closest to 299 bp with smallest standard deviation for sequencing

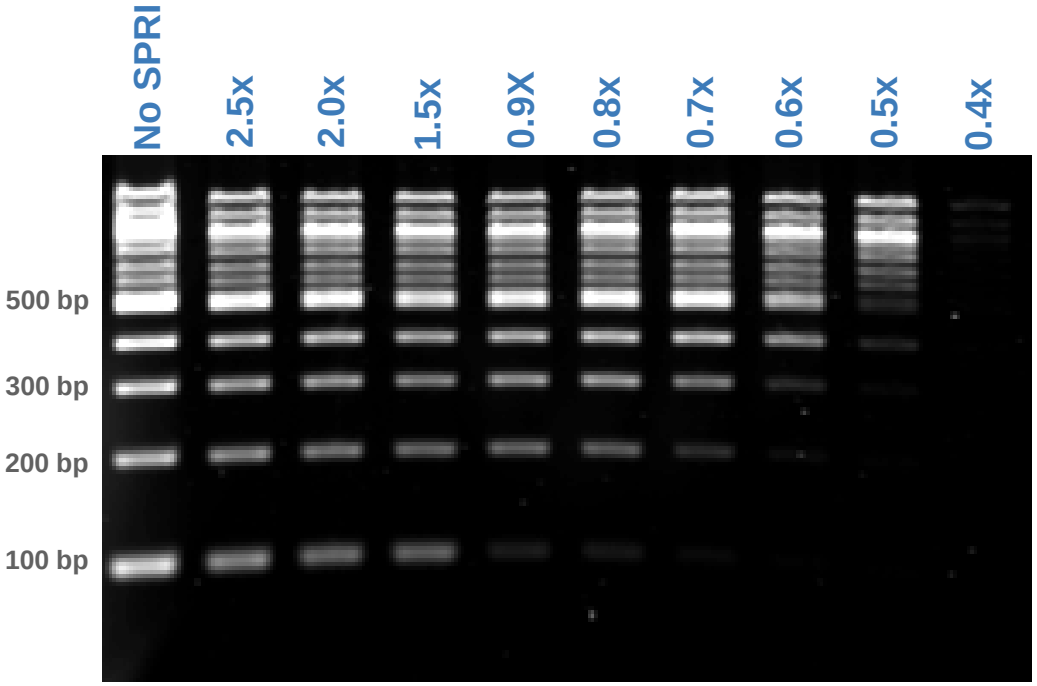




BROAD INSTITUTE illumina

Option 2: Double-Sided SPRI

- As discussed previously, the size of fragments that SPRI binds can be varied by varying the “X” fo SPRI that is added to the reaction containing the DNA:



- We can utilize this property to perform a gel-free and automated size selection that we have termed “Double-sided SPRI”

BROAD INSTITUTE illumina

Option 2: Double-Sided SPRI

- By implementing a combination of good shearing with SPRI and “reverse” SPRI, one can select a fairly tight size range *with no gel*:

0.65X SPRI

Fragments >250 bp on beads
Fragments <250 bp in supernatant

KEEP SUPERNATANT
0-250 bp

→

DISCARD BEADS
“Reverse SPRI”

1.0X SPRI

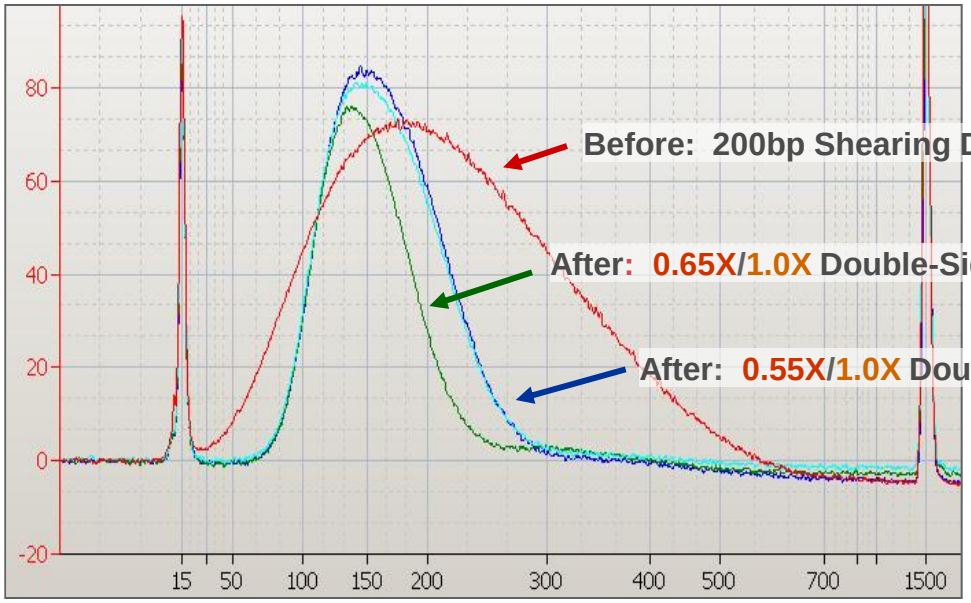
Fragments >100 bp on beads
Fragments <100 bp in supernatant

KEEP BEADS
100-250 bp

→

DISCARD SUPERNATANT

Results:

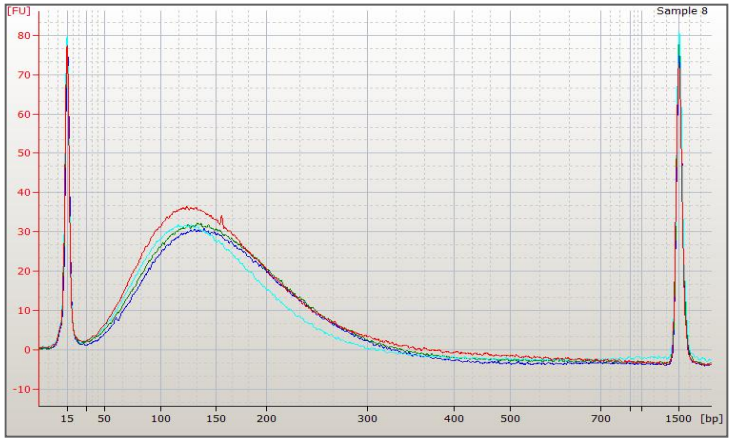


BROAD INSTITUTE illumina

Option 3: No Size Selection At All

- If tight sizing not required for a project, size selection may be unnecessary.
- Combining a tight 150 bp shear with a 0.7X SPRI clean-up after adapter ligation to remove excess adapter dimer may be suitable.

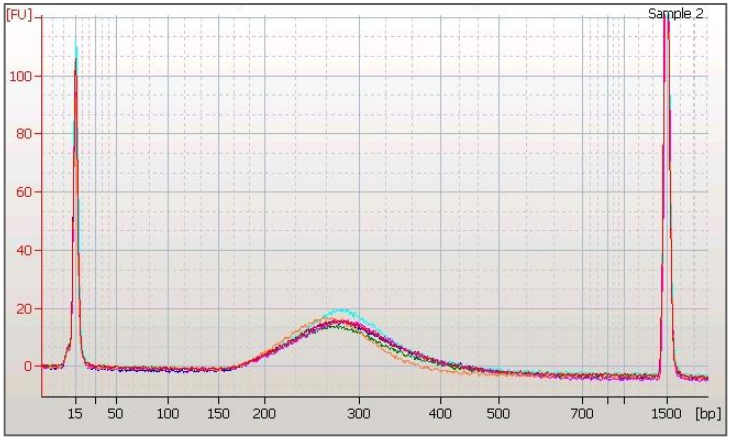
POST SHEARING (150 bp protocol)



End Repair
A Addition
Adapter Lig
NO GEL
O.7X SPRI

PCR Enrichment
1 ul DNA
10 cycles

POST PCR ENRICHMENT & 0.7X SPRI

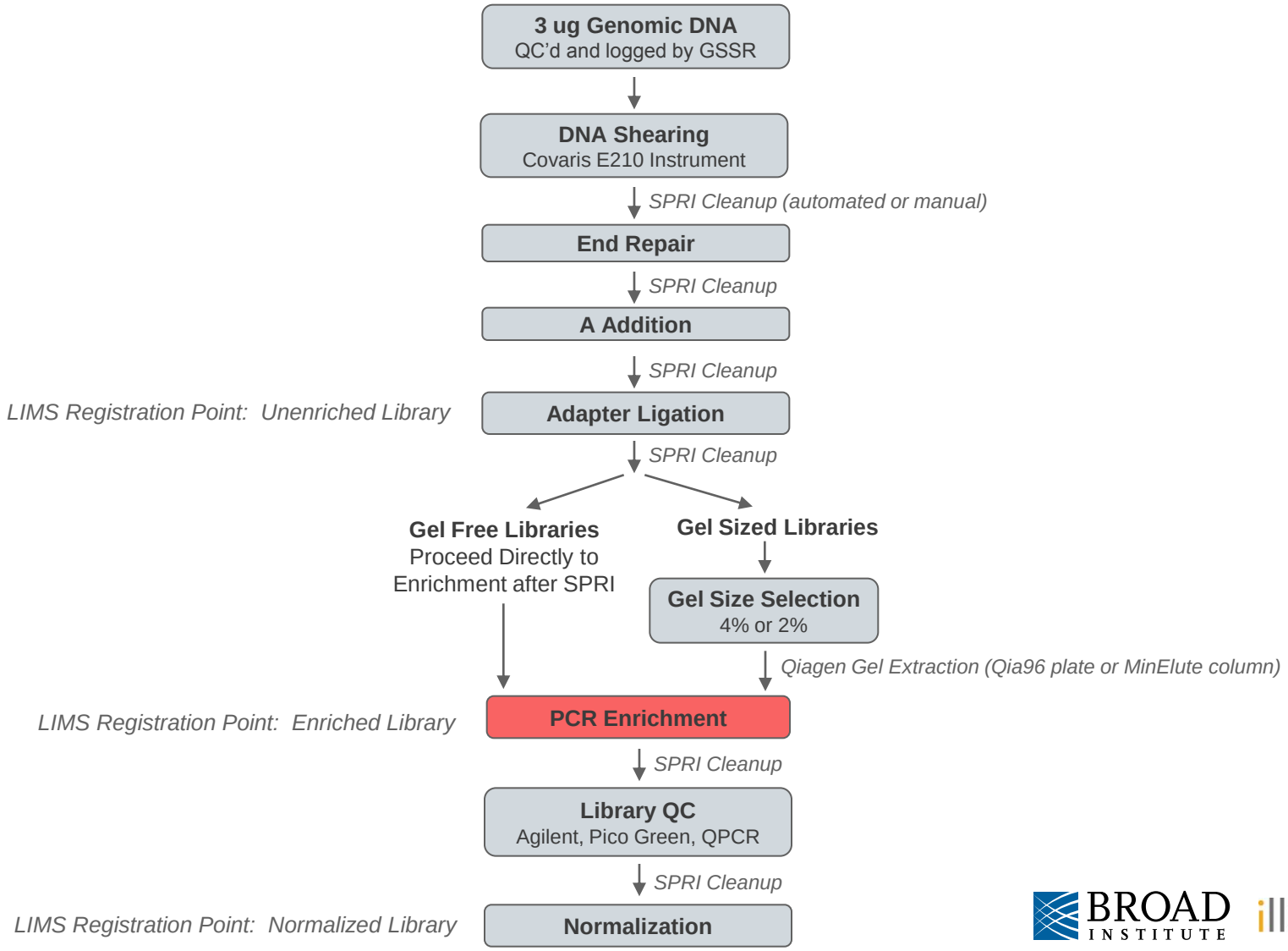


- This works because of the tightness of our 150 bp shear. If shearing isn't optimized or if a larger size is needed, than double-sided SPRI is required.

BROAD INSTITUTE illumina

PCR Enrichment

PCR Enrichment



Library Complexity

- Library complexity refers to the number of unique fragments present in a given library
- Complexity is affected by:
  - Amount of starting material
  - Amount of DNA lost during cleanups and size selection
  - Amount of duplication introduced via PCR
- For most libraries that only need to be run across a few lanes, the standard protocol provides libraries with ample complexity
- However, certain projects require very deep coverage from a single sample - i.e. SNP discovery, mammalian assembly, cancer resequencing
- When dozens of lanes are required, library complexity becomes very important

BROAD INSTITUTE illumina

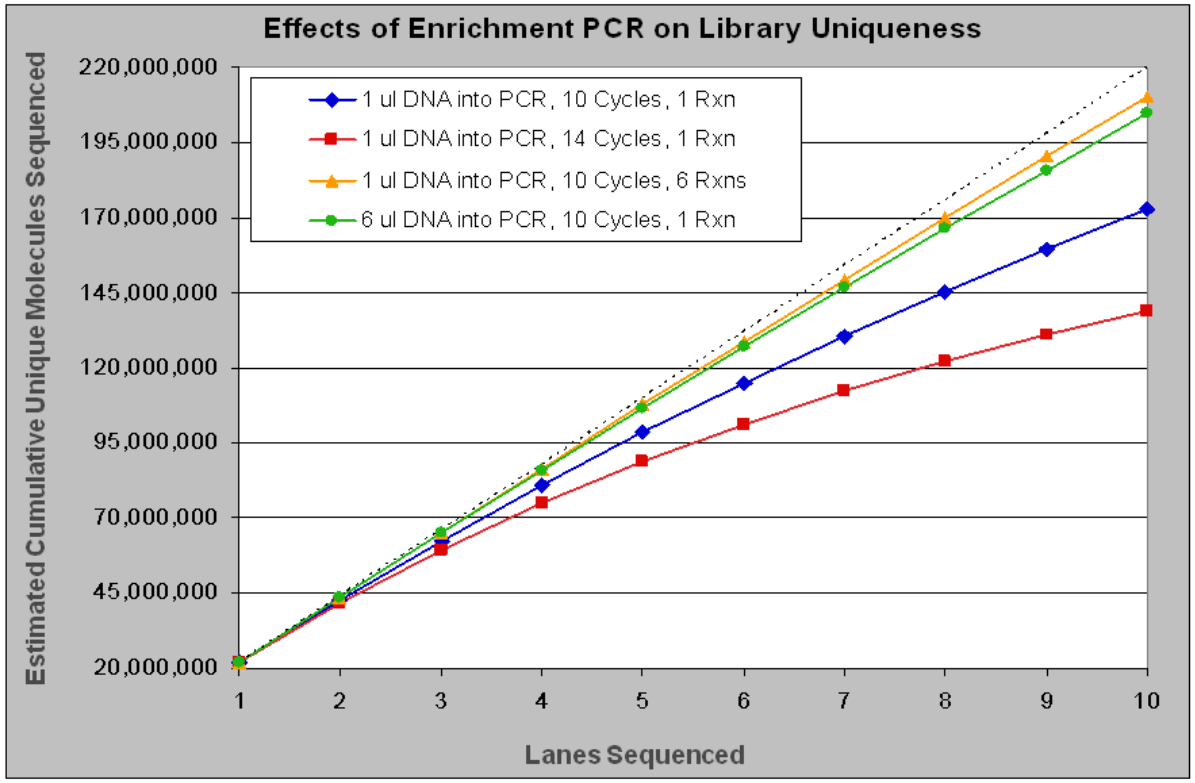
Relationship Between PCR and Complexity

Experiment:

- Varied input volume, PCR cycle number, & number of reactions pooled
- Sequenced to determine % duplication & estimated library size
- Based on data, number of new molecules seen after each successive lane of sequencing was estimated

Results:

- Decreasing PCR cycles decreases % duplication
- Increasing DNA input and pooling multiple PCR reactions decreases % duplication



Starting DNA	Cleanup Method	DNA Input to PCR (ul)	PCR Cycle #	# of PCR Rxn Pooled	% Duplication	Estimated Library Size
3 ug	MinElute	1	10	1	2.79%	247,441,814
3 ug	MinElute	1	14	1	5.51%	195,858,361
3 ug	MinElute	6	10	1	0.83%	1,255,845,288
3 ug	MinElute	1	10	6	0.48%	1,390,926,598

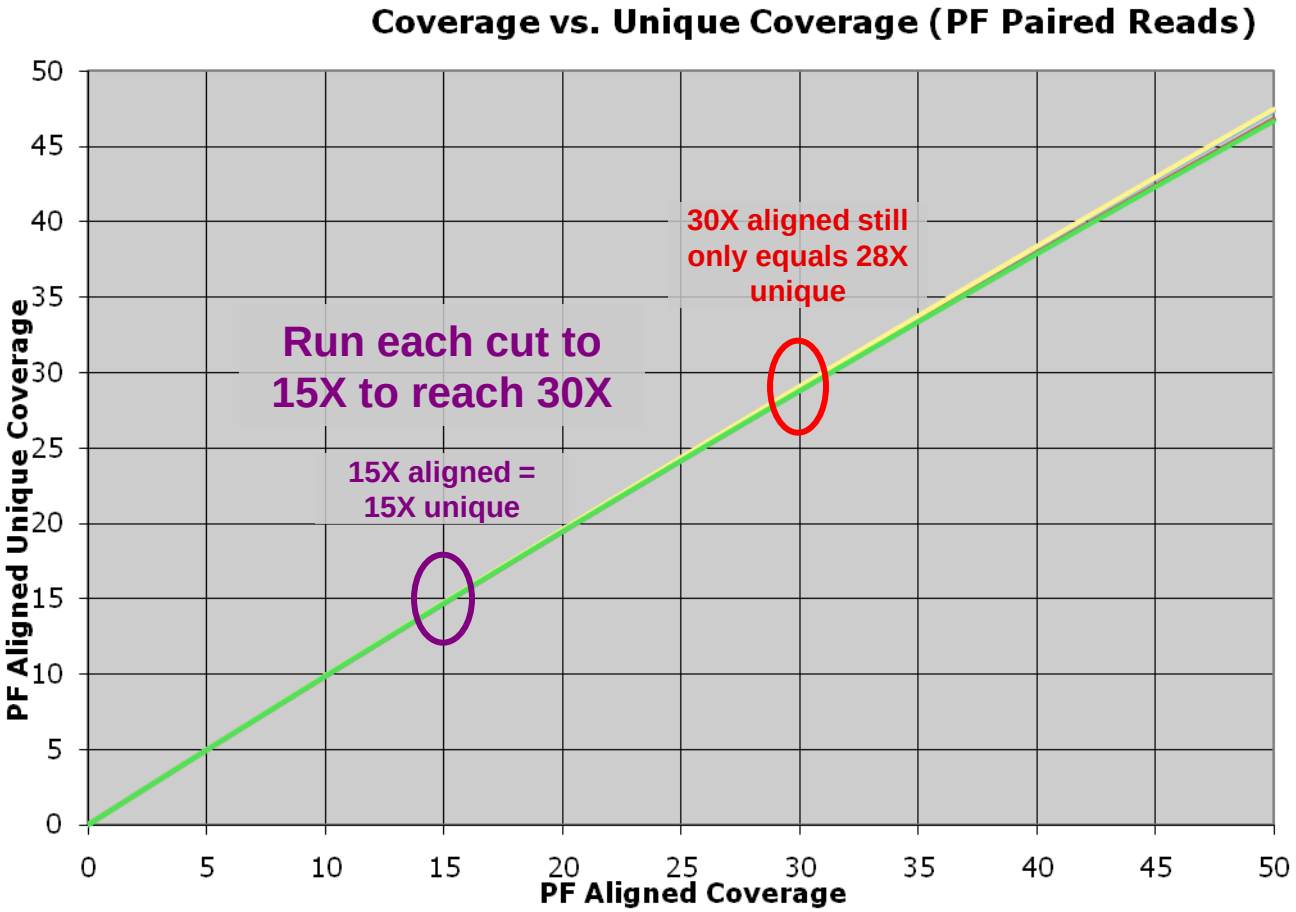
BROAD INSTITUTE illumina

Application - “High Diversity” Sample Preparation

- Cancer resequencing projects required 30X unique coverage from tumors and their matched normals
- We needed a way to reach this goal while minimizing both the amount of genomic DNA used and the amount of sequencing lanes needed
  - Highest complexity, lowest % duplication
  - Maintain a very tight size distribution of about 500 bp +/- 10%
- To accomplish this, we developed a high diversity protocol:
  - Take **two** extremely tight gel cuts from each sample: 500bp & 520bp
    - Creates two libraries for each 3 ug input
    - Double the amount of fragments from the same starting input
  - Enrich **ENTIRE** sample from both cuts to maximize diversity
    - 16 x 50 ul PCR reactions with 10 cycles enrichment
    - Pool 16 reactions for each cut & clean up in same column or SPRI

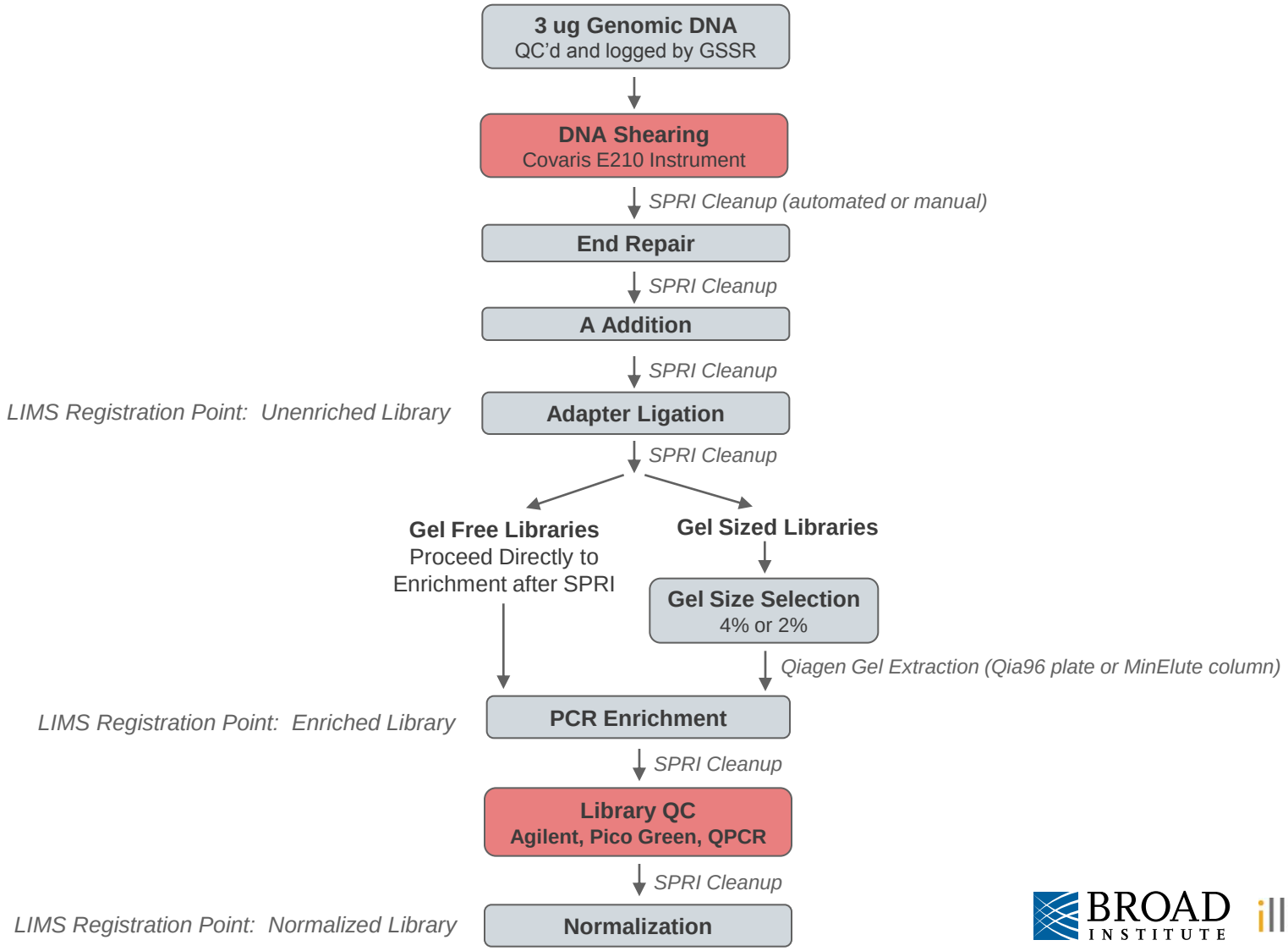
BROAD INSTITUTE illumina

Application - “High Diversity” Sample Preparation



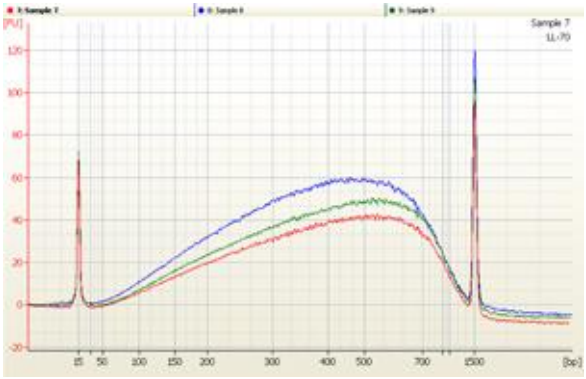
Library Quality Control

Library Quality Control



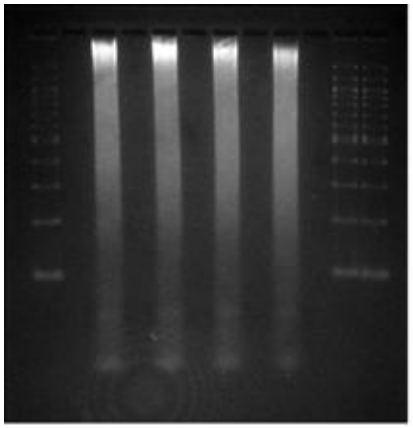
Library Quality Control

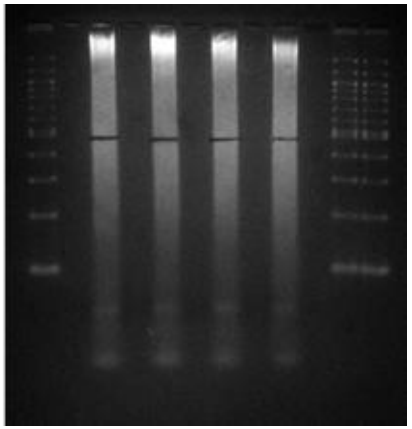
- **Agilent BioAnalyzer 2100**
  - Ensure proper size distribution after both shearing & PCR
  - Failing samples enter rework loop: re-sheared or re-enriched



- **Pico Quantification Post-Enrichment**
  - Samples <20nM are re-enriched with more input and/or cycles

- **Gel Images**
  - Look for excessive adapter, DNA in unexpected size range, etc.
  - Pictures taken before and after cutting are saved for downstream troubleshooting





BROAD INSTITUTE illumina

Library Quality Sequencing Metrics

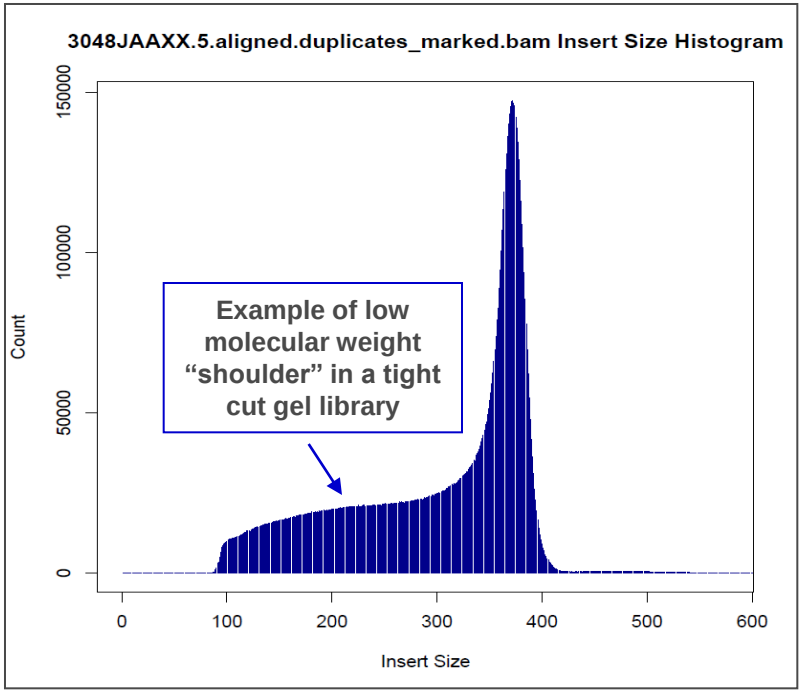
- **Always check the sequencing metrics for a run before looking at the LC metrics**

Metric	Definition	How to Interpret…
% Aligned	% of passing filter (PF) reads that aligned to reference sequence.	Low % aligned may indicate sample contamination or swap.
% Adapter	% of PF reads whose first 16 bases match any part of the Illumina adapter sequence.	>1% adapter indicates inefficient removal of adapter dimer in size selection.
% Chimerism	% PF of reads that have 2 ends over 100kb apart or on 2 chromosomes.	>1% chimerism indicative of problem in adapter ligation or with genomic DNA prep itself.
% Duplication	% of PF aligned reads originating from duplicate fragments (i.e. multiple reads with exact same R1 and R2 start sites).	High % duplication indicates a low complexity library, possibly due to low amount of starting material and/or excessive PCR cycles.
Estimated Library Size	Estimated number of unique molecules in library, calculated using % duplication and reference genome.	Library size depends on starting material. Human WGS libraries should have a size of 1-3 billion.

BROAD INSTITUTE illumina

Library Quality Sequencing Metrics

Metric	Definition	How to Use Interpret…
Concordance	Concordance of sequencing results with previous genotyping/SNP data.	Can detect sample swap or contamination events.
Insert Size Plots	Show the size distribution of sequenced fragments.	“Shoulders” or spurious peaks indicate problem in size selection.



BROAD INSTITUTE illumina

Module Summary - Sample Preparation

- Sample preparation is the most customizable area of the Illumina process - Can adapt the basic protocol to meet the goals of any project
- Sample tracking is important for any size laboratory to ensure sample integrity
  - High throughput: barcoded libraries, plates, and LIMS tracking
  - Low throughput: Excel sheets or databases to record project details & QC information
- Broad has developed many modifications to the standard protocol to increase throughput and reproducibility, as well as to meet specific project needs:
  - Optimized shearing with Covaris
  - Implementation of AMPure SPRI for clean-ups and gel-free size selection
  - Optimized gel cutting conditions for libraries with <10% size deviation
  - Development of high diversity protocols for deep coverage libraries
- Library quality control monitoring and sample tracking are essential to ensure that only high quality libraries make it to the sequencer

BROAD INSTITUTE illumina

What You Will See in the Lab Sections:

- In lab sample tracking methods and tools
- Covaris E210 demo
- Agarose gel cutting tips & tricks
- Library QC methods and review of example data
- SPRI clean-up hands-on demo & automation overview

BROAD INSTITUTE illumina