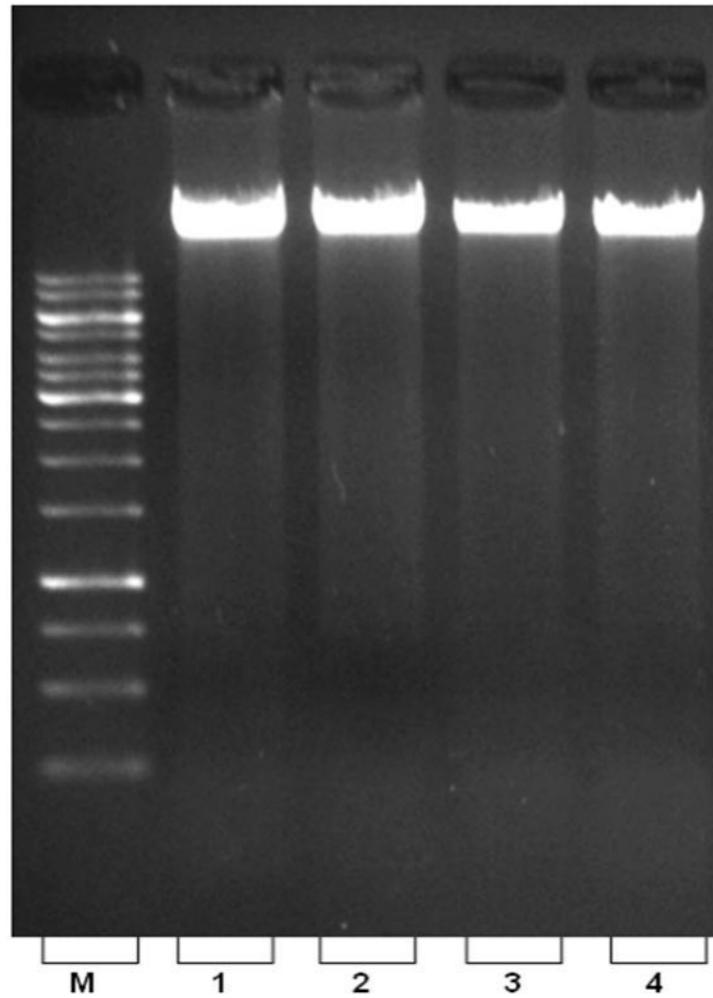


# Librerías o genotecas

Genómica para bioinformática INB320  
3 de agosto de 2017  
Eduardo Castro, PhD  
[www.castrolab.org](http://www.castrolab.org)

# ¿Cómo pasamos de DNA a secuencias?



ATGG**C**ATTGCAA  
TGG**C**ATTGCAATTG  
AGATGG**T**ATTG  
GATGG**C**ATTGCAA  
GCATTGCAATTGAC  
ATGG**C**ATTGCAATT  
AGATGG**T**ATTGCAATTG  
  
AGATGG**C**ATTGCAATTGAC

¿Por qué un bioinformático debería saber cómo preparar librerías?

# ¿Por qué un bioinformático debería saber cómo preparar librerías?

- Probablemente no les interesa

# ¿Por qué un bioinformático debería saber cómo preparar librerías?

- Probablemente no les interesa
- Probablemente eso fue lo que los llevó a la bioinformática

# ¿Por qué un bioinformático debería saber cómo preparar librerías?

- Probablemente no les interesa
- Probablemente eso fue lo que los llevó a la bioinformática
- **El análisis de secuencias depende de cómo se preparó la librería**

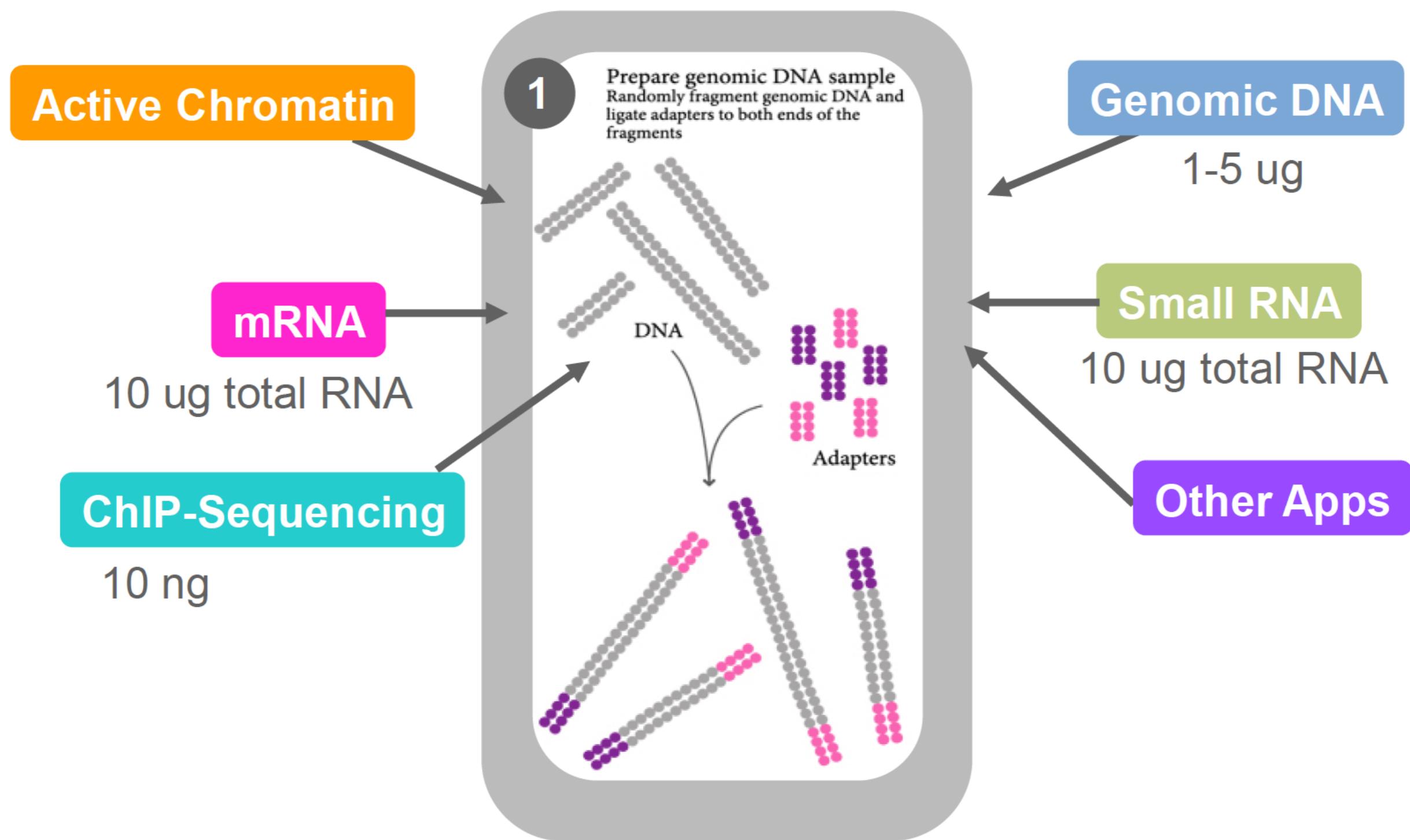
# ¿Por qué un bioinformático debería saber cómo preparar librerías?

- Probablemente no les interesa
- Probablemente eso fue lo que los llevó a la bioinformática
- **El análisis de secuencias depende de cómo se preparó la librería**
- **Lengua común entre investigadores**

# For all you seq



# Preparación de librería: ligar adaptadores a fragmentos de DNA



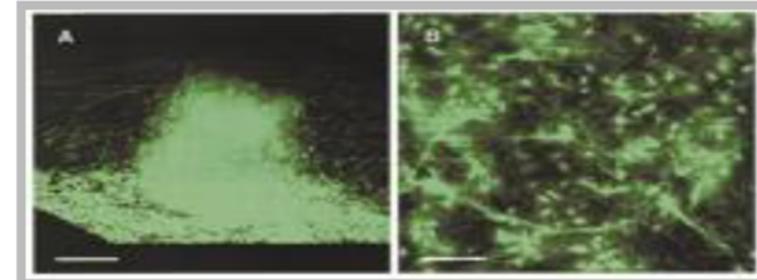
# Preparación de librería: ligar adaptadores a fragmentos de DNA

- Se requiere conocer la cantidad de material inicial con que se cuenta ya que distintas aplicaciones tienen distintos requerimientos

## Methods for Quantification:



NanoDrop\*



PicoGreen® \*

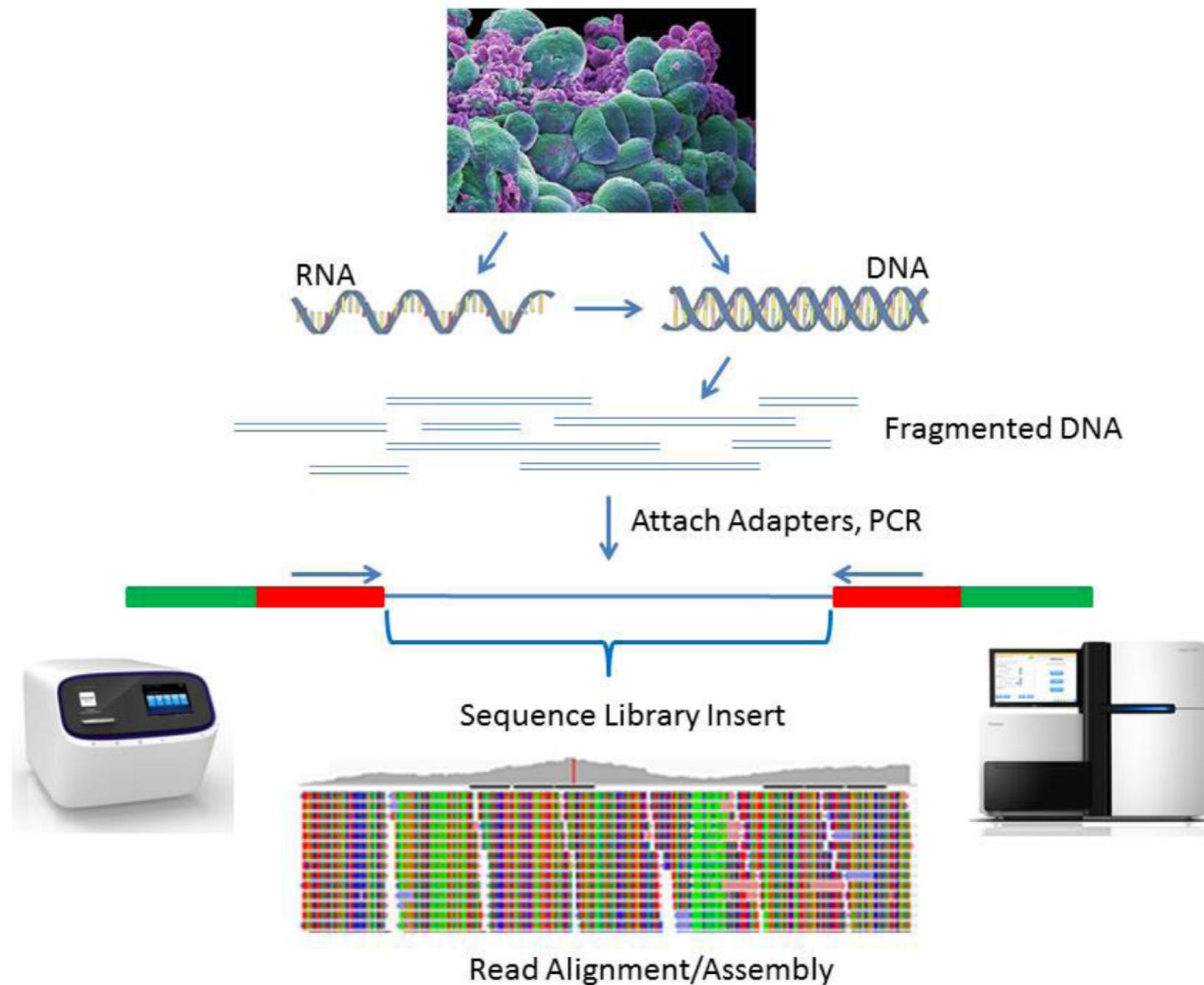


Qubit™

# Todo está en la preparación de la librería

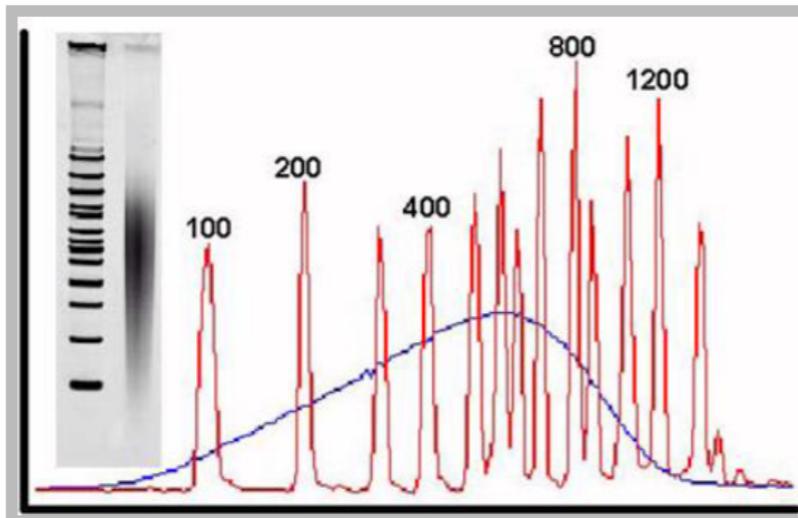
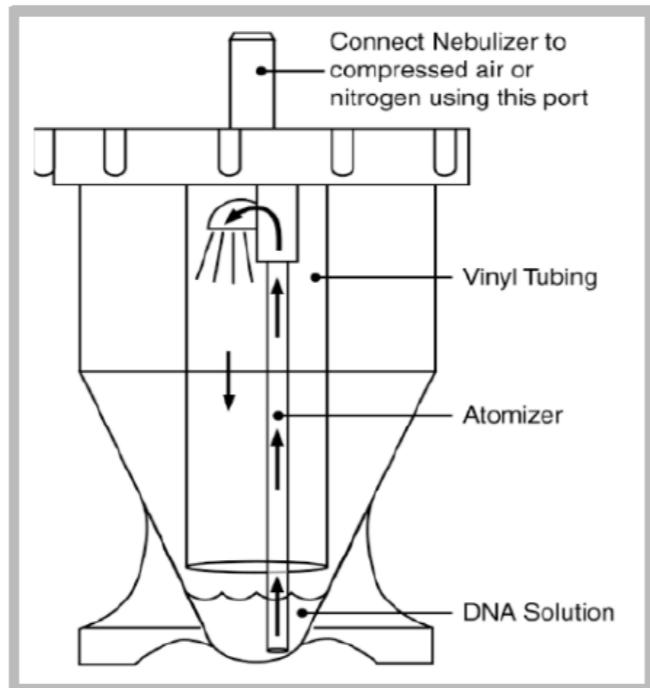
1. Fragmentación
2. Selección del tamaño de inserto
3. Construcción de la librería
4. Control de calidad

# Todo está en la preparación de la librería



# Fragmentación de ácidos nucleicos

# Fragmentación enzimática o física



## ► Nebulizer

- Very inexpensive
- Works well for 1-5 µg starting input

## ► Other methods available

- Covaris™
- Sonication
- HydroShear®
- Enzymatic
- Others

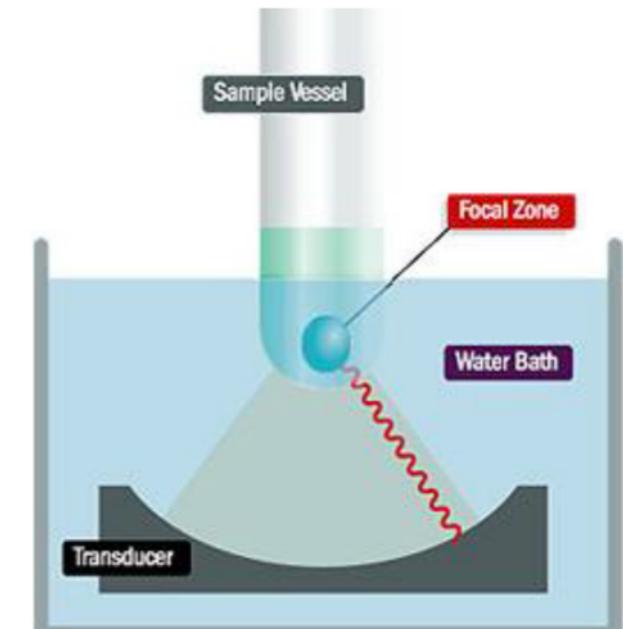
# Fragmentación física

- Covaris → ultrasonido
- Variando la frecuencia y tiempo de las ondas de sonido se obtienen fragmentos de diferentes tamaños. 100 bp-10kb
- No-contacto, tubo cerrado, isotérmico
- También 6-20 kb para mate-pairs



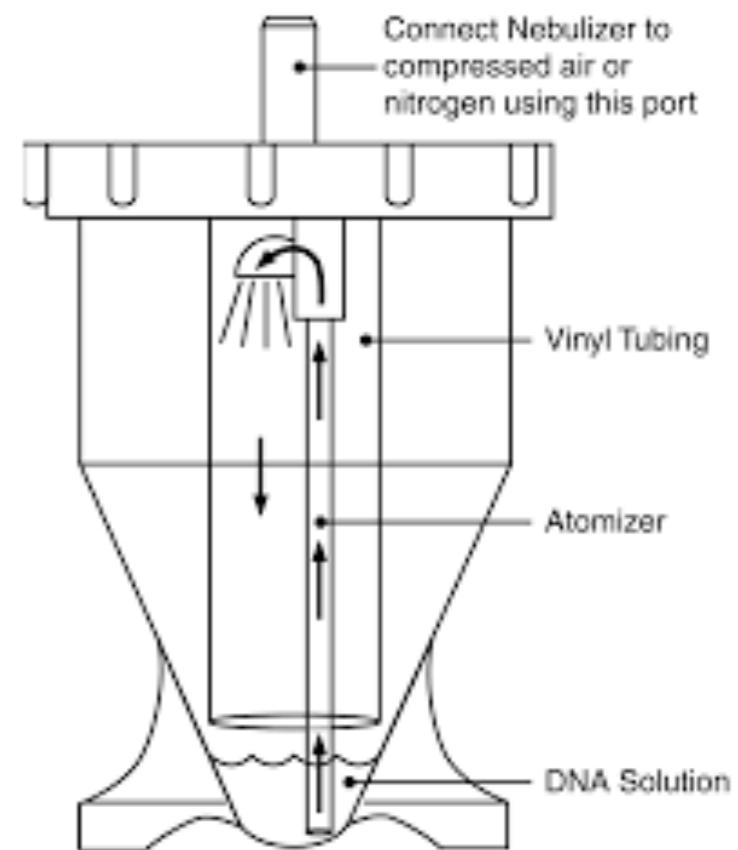
M220 Focused-ultrasonicator™

DNA Shearing for Next-Generation Sequencing



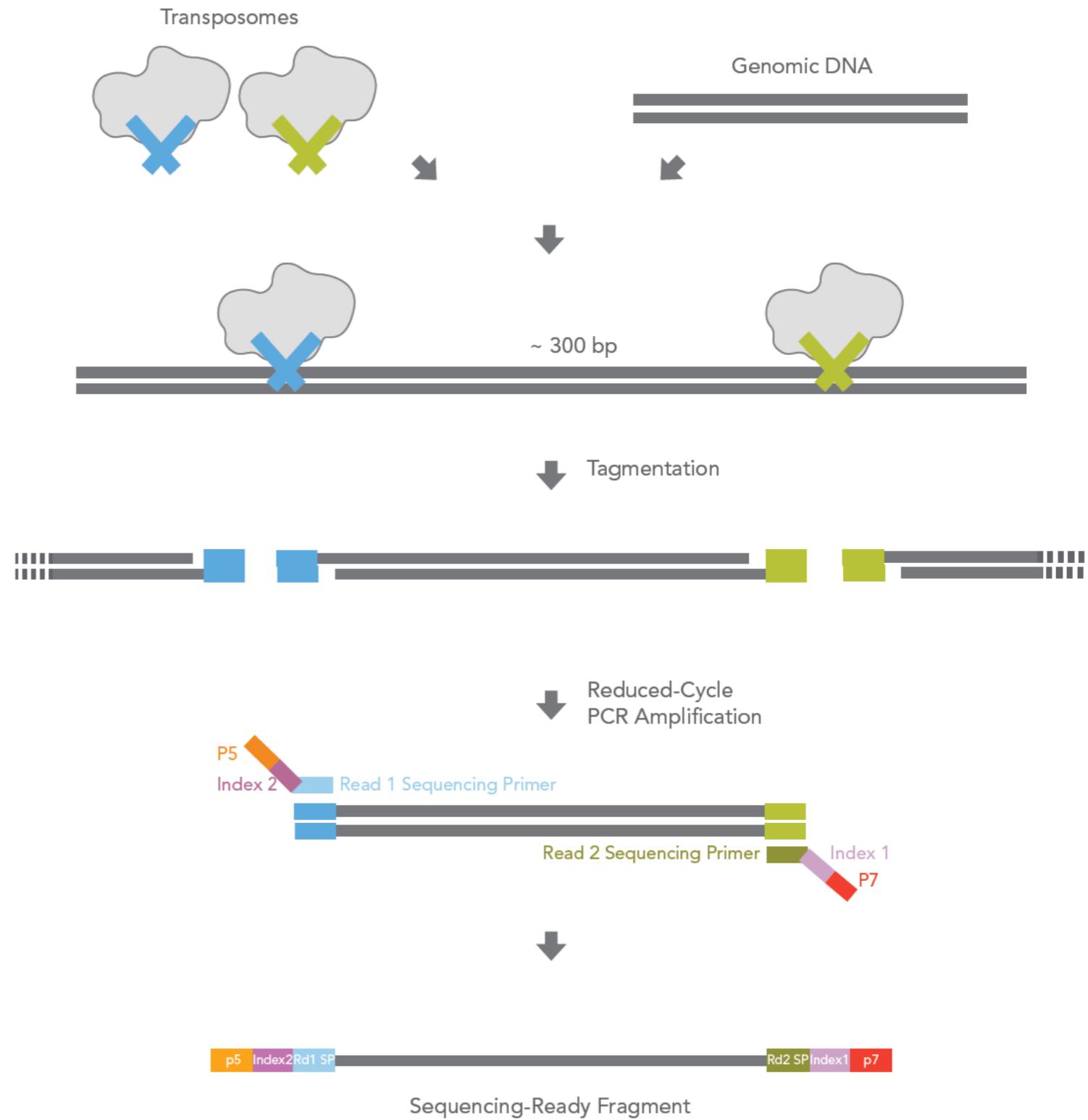
# Fragmentación física

- Nebulizer → atomización
- Atomizar líquido por presión de aire u otro gas, e.g., N<sub>2</sub>
- 100 bp - 3 kb
- No se recomienda si tienes poco material de entrada.  
Se pierde 30% de la muestra



# Fragmentación enzimática

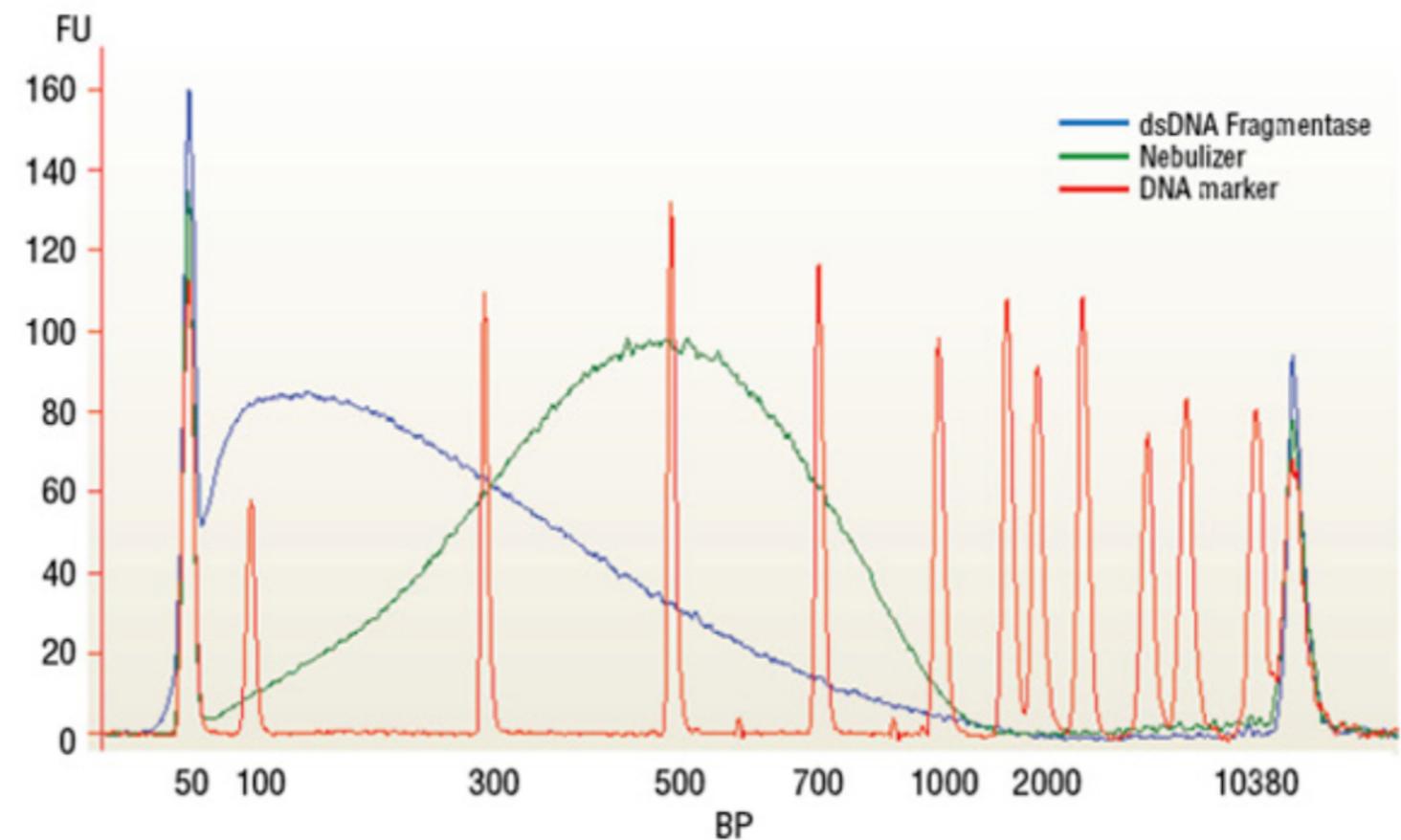
- Transposasa
- Nextera toma 90 minutos
- 50 ng input
- Inserto ~250 bp



# Fragmentación enzimática

- Fragmentasa
- NEB
- Nucleasa mutante de *Vibrio vulnificus* una endonuclease mutante de fago T7
- Rápido

NEBNext dsDNAFragmentase® generates fragments in the 100–300 bp range more effectively than nebulization.



# La fragmentación introduce sesgos en el secuenciamiento

- Válido para fragmentación enzimática y física
- Impacto en “genome coverage”
- Probabilidad de secuenciar una región desfavorecida

Article | OPEN

## Non-random DNA fragmentation in next-generation sequencing

Maria S. Poptsova, Irina A. Il'icheva, Dmitry Yu. Nechipurenko, Larisa A. Panchenko, Mingian V. Khodikov, Nina Y. Oparina, Robert V. Polozov, Yury D. Nechipurenko & Sergei L. Grokhovsky ✉

*Scientific Reports* **4**, Article number: 4532  
(2014)

doi:10.1038/srep04532

[Download Citation](#)

Bioinformatics Biopolymers in vivo

Received: 19 September 2013

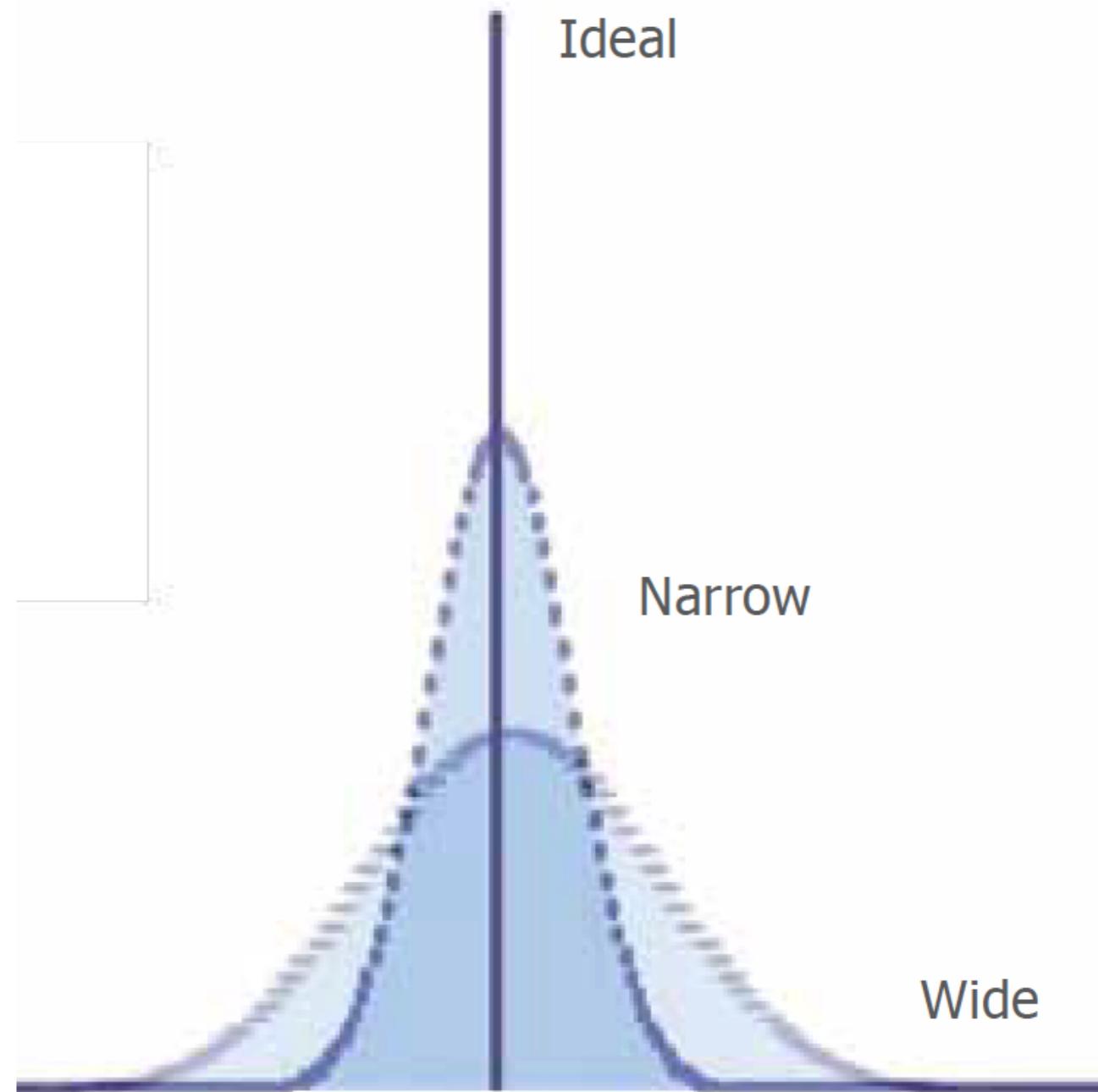
Accepted: 13 March 2014

Published online: 31 March 2014

# Selección del tamaño de inserto

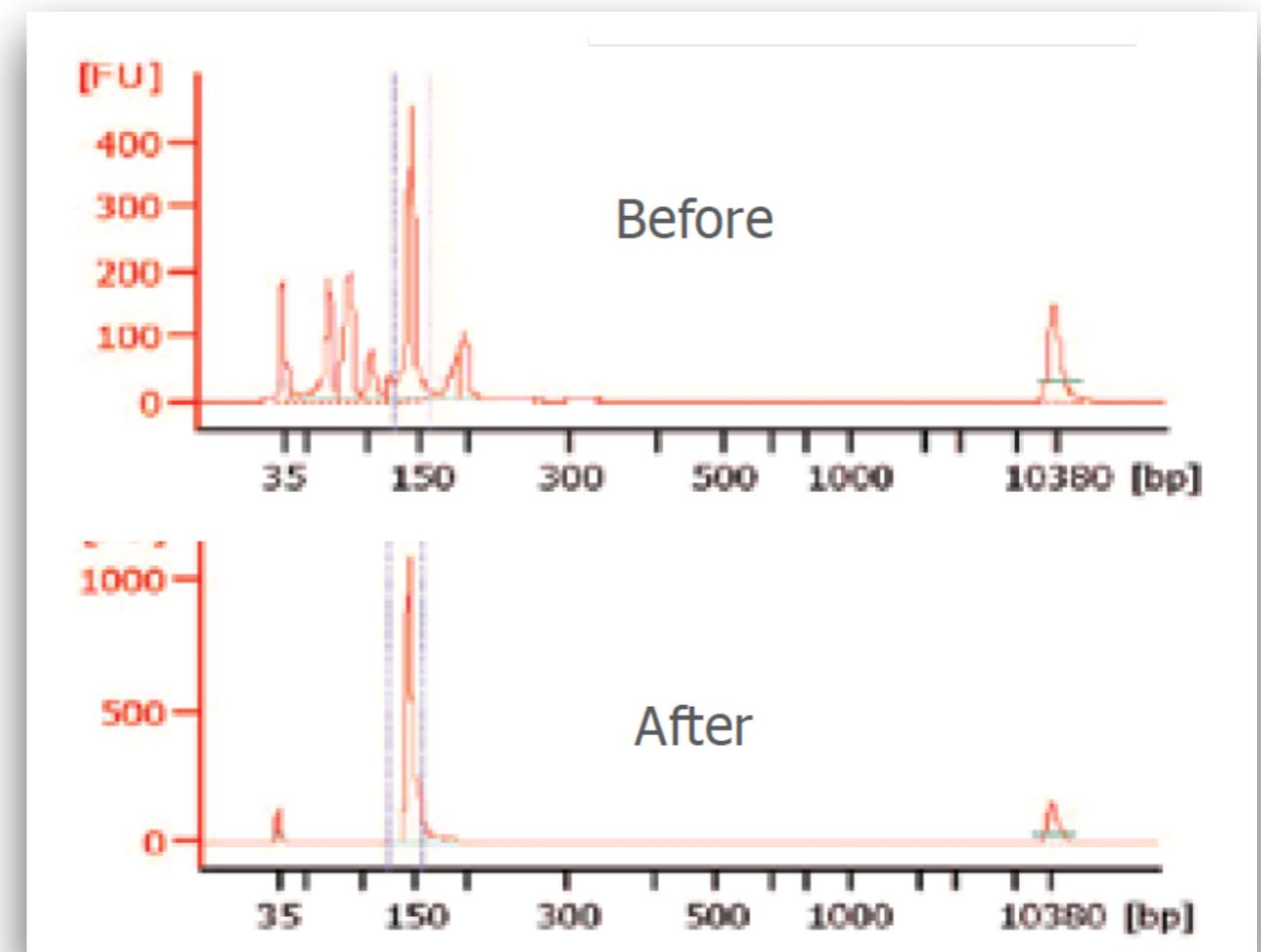
# Distribución de fragmentos

- Fragmentación produce una distribución amplia de fragmentos
- Mayor eficiencia de secuenciación cuando la distribución es más ajustada



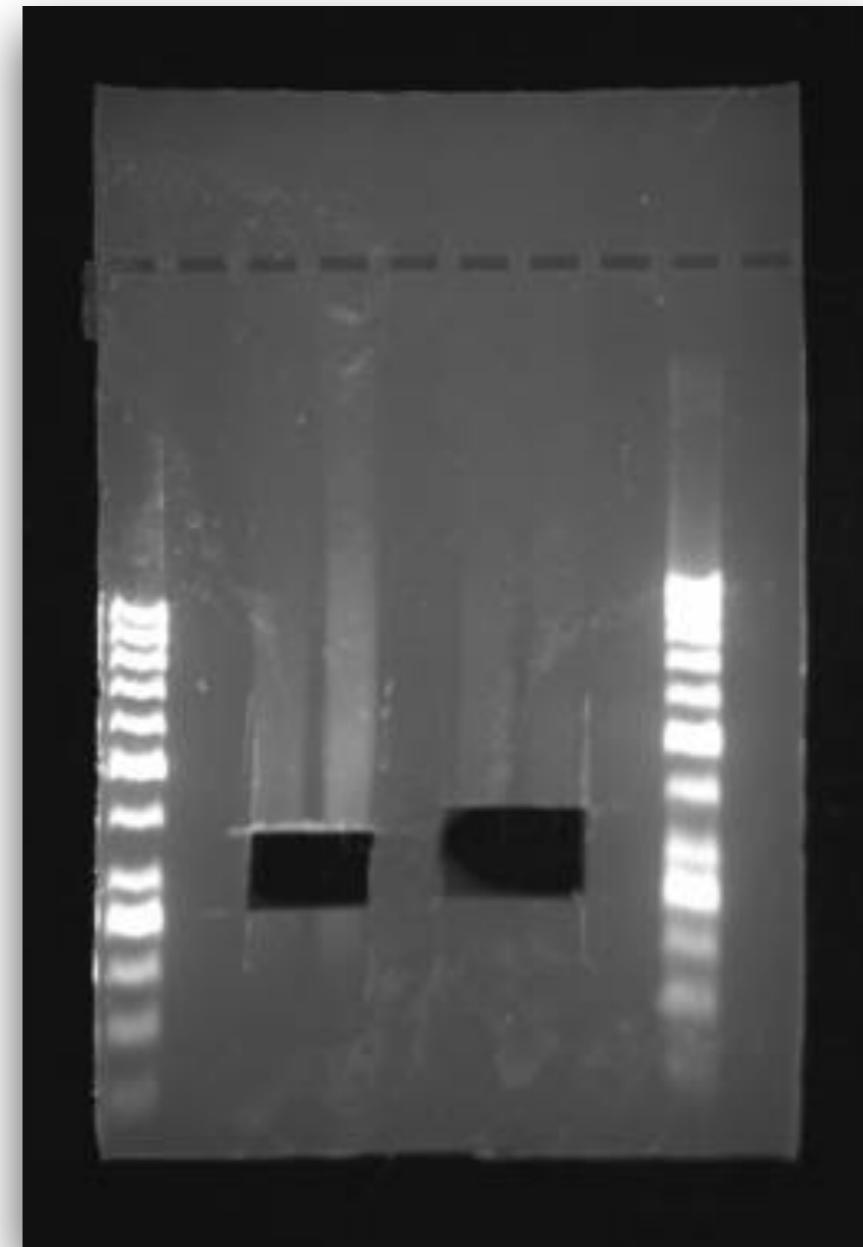
# Seleccionar solo fragmentos del tamaño adecuado

- Illumina está optimizado para 400-600 bp
- Gel
- Beads o “perlas”
- Columnas
- Cassettes



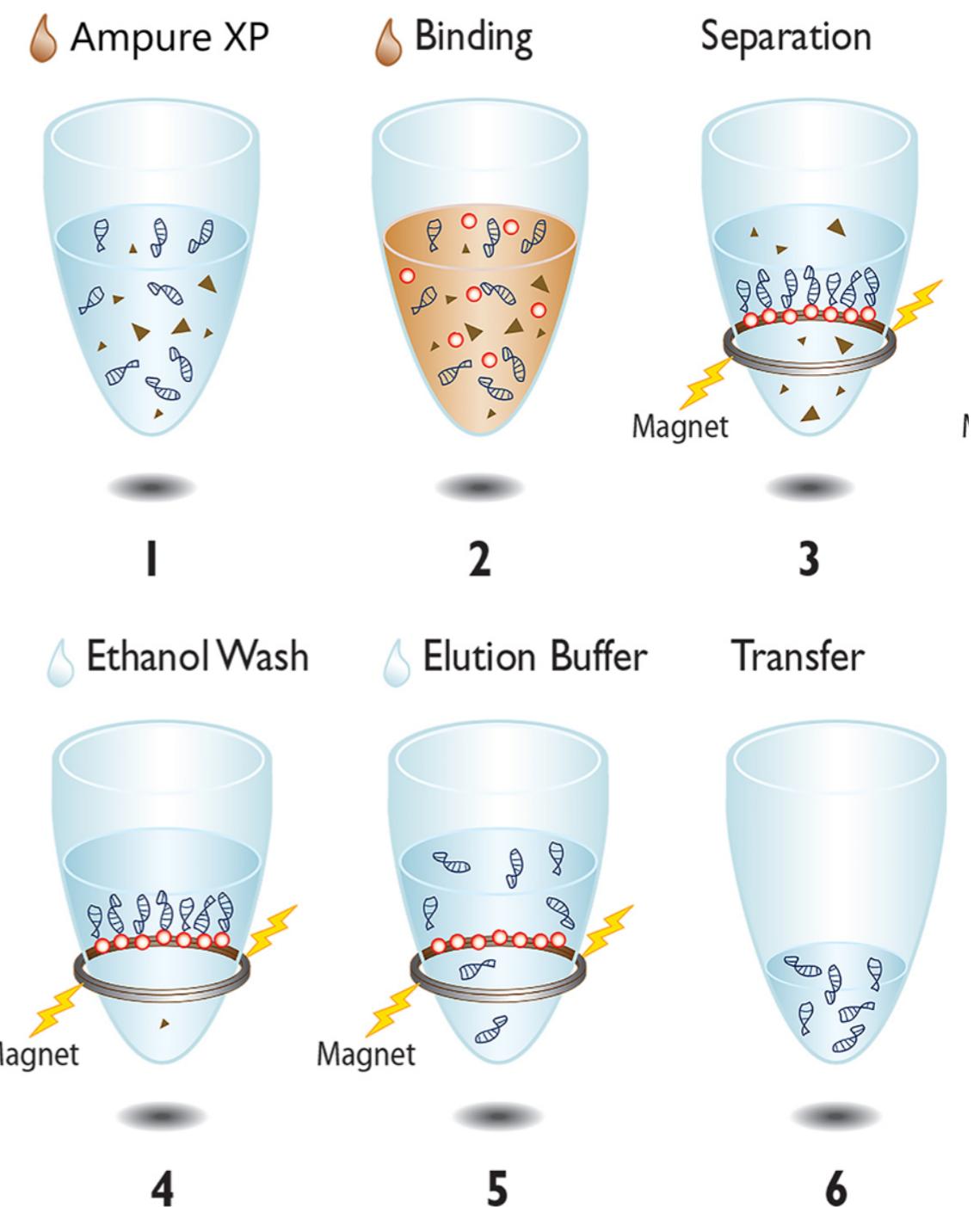
# Selección de tamaño de inserto

- Gel
  - Laborioso
  - No es automatizable
  - No es compatible con procesar muchas muestras
  - Radiación UV
  - Distribución de fragmentos resultante no es tan ajustada



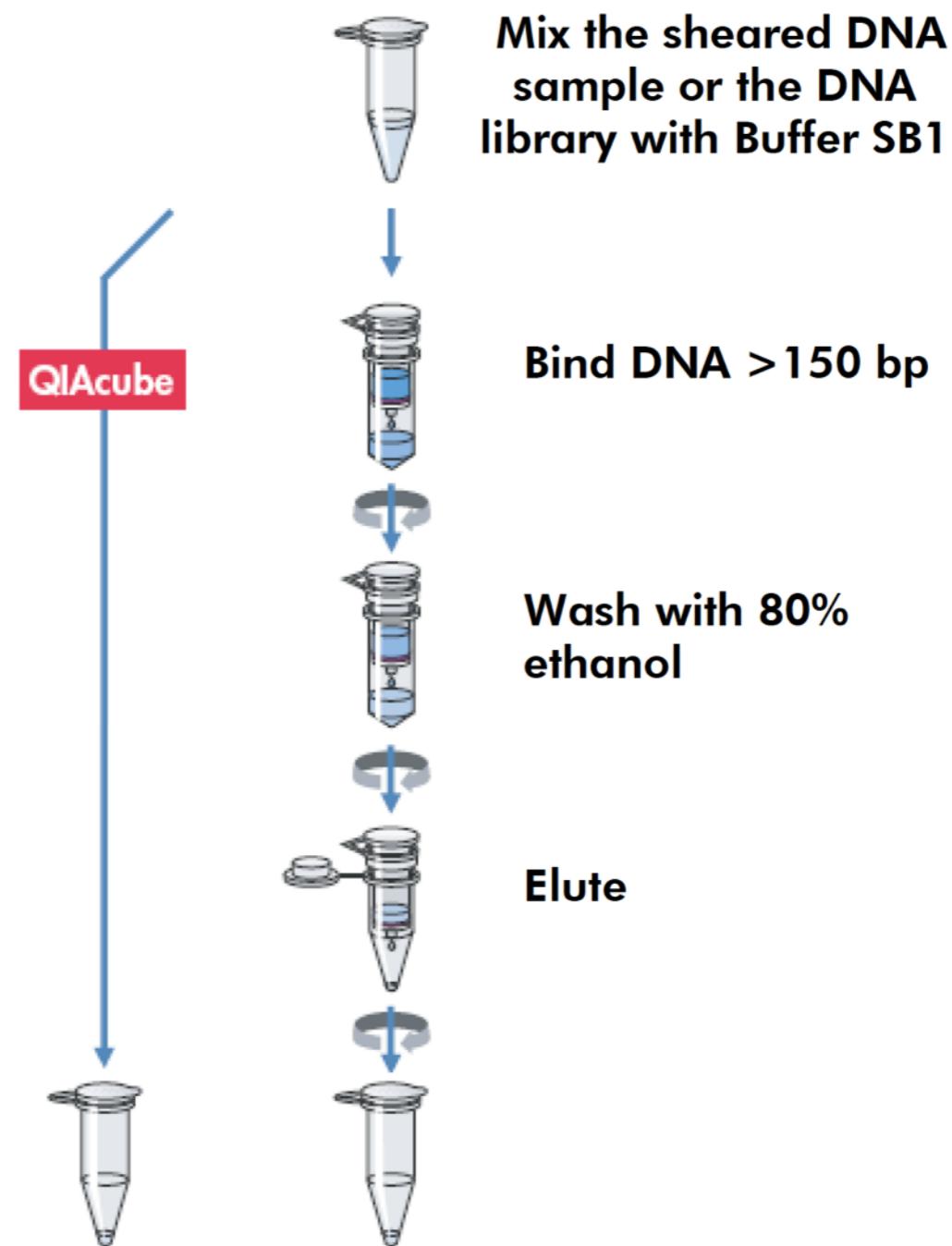
# Selección de tamaño de inserto

- Beads o “perlas”
- Solid Phase Reversible Immobilisation (SPRI) beads
- Muy bueno pero caro
- Se puede automatizar
- No es peligroso
- Rápido



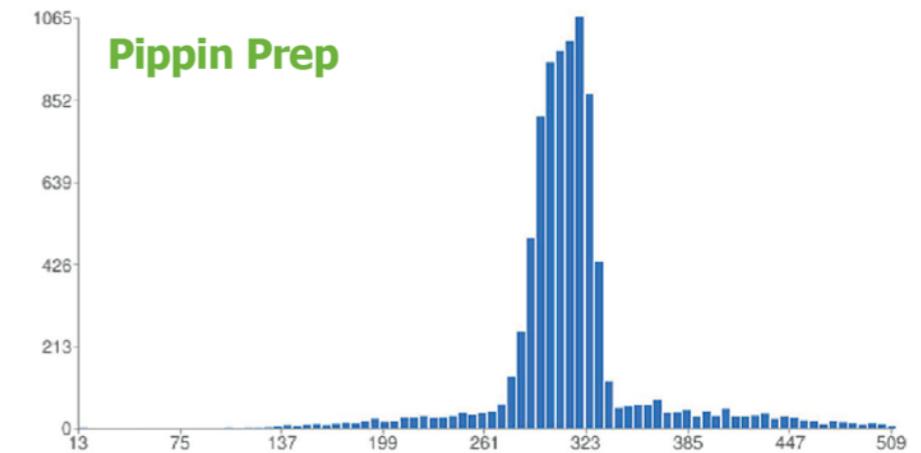
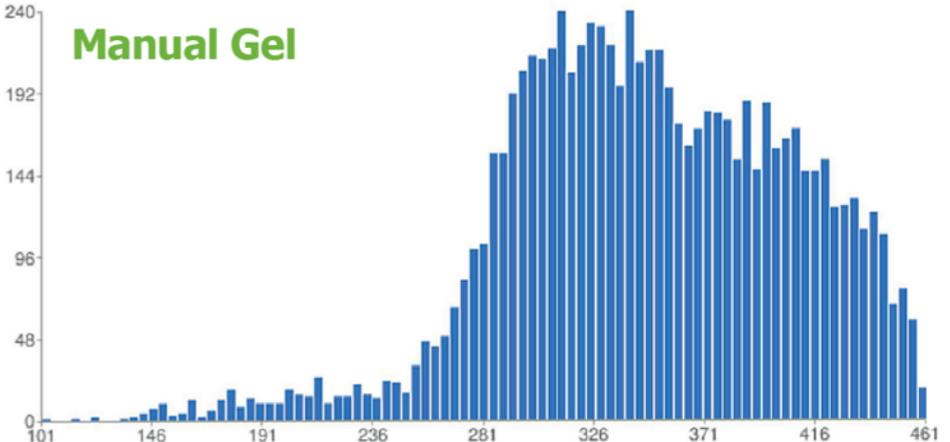
# Selección de tamaño de inserto

- Columnas
- Similar a una extracción de ácidos nucleicos
- Basado en silica
- > 150 bp
- Automatizable



# Selección de tamaño de inserto

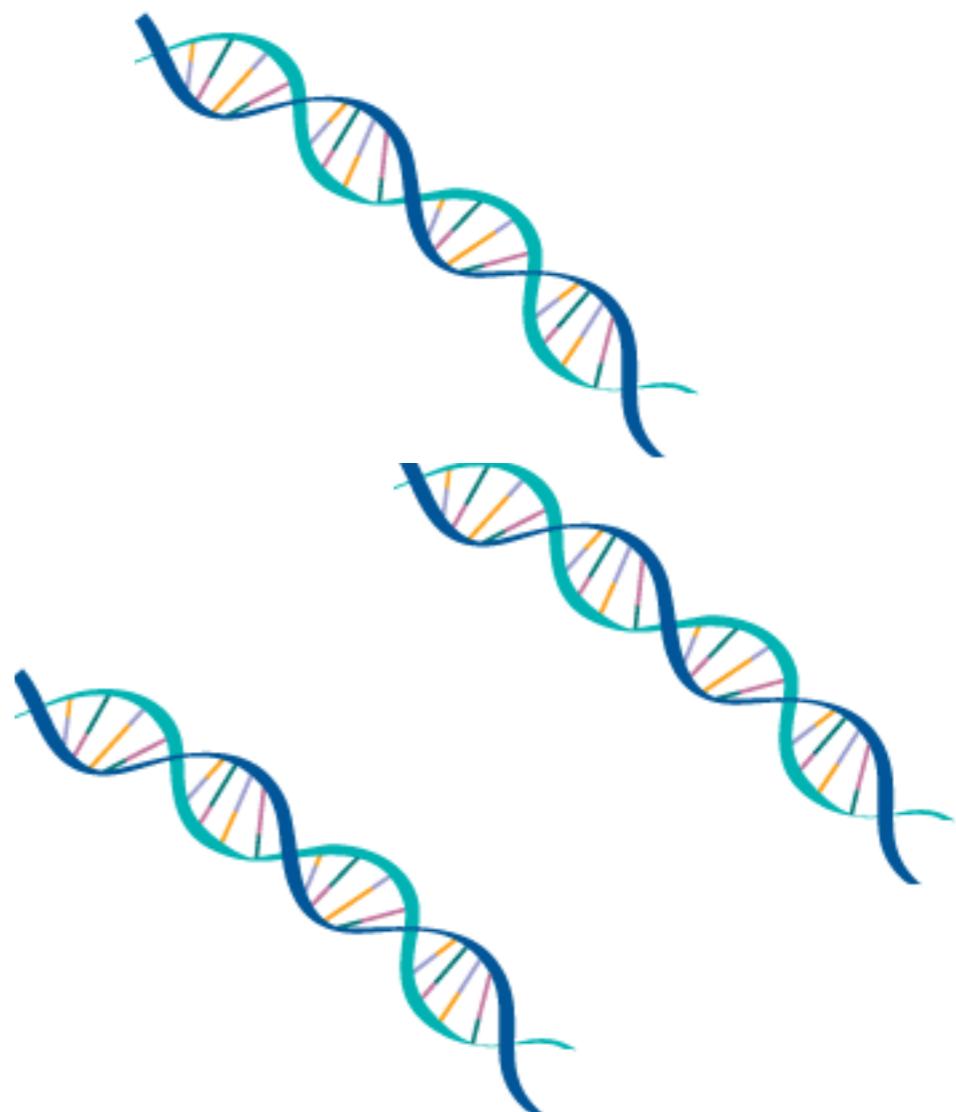
- Cassettes o geles pre empaquetados
- Buenos resultados
- 90 bp - 50 kb
- 24 muestras a la vez
- Reproducible



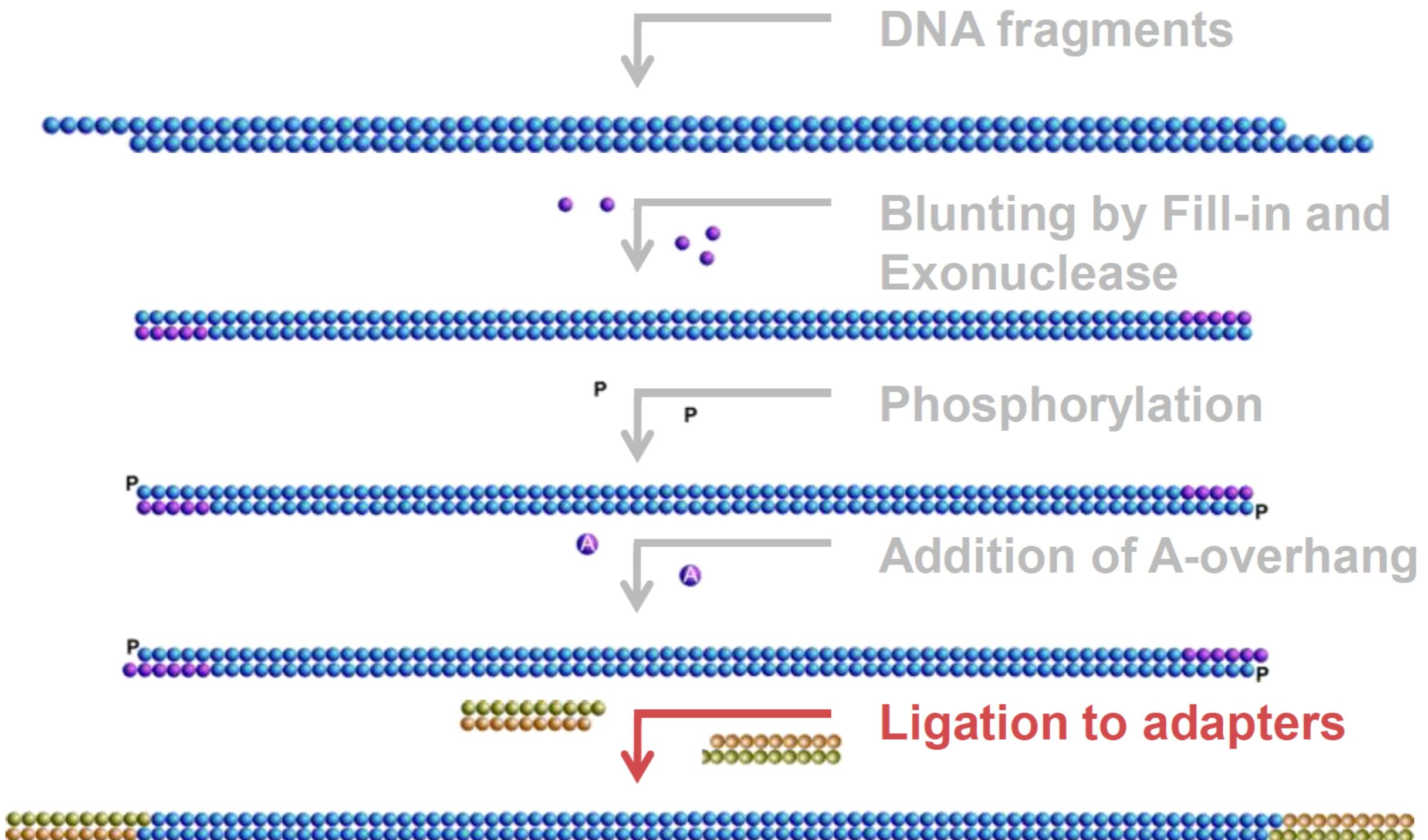
# Construcción de la Librería

# ¿Qué tenemos hasta ahora?

- Un montón de fragmentos de DNA
- Ojalá más o menos del mismo tamaño
- ¿Qué es lo que necesitamos?



# ¿Qué es lo que necesitamos?

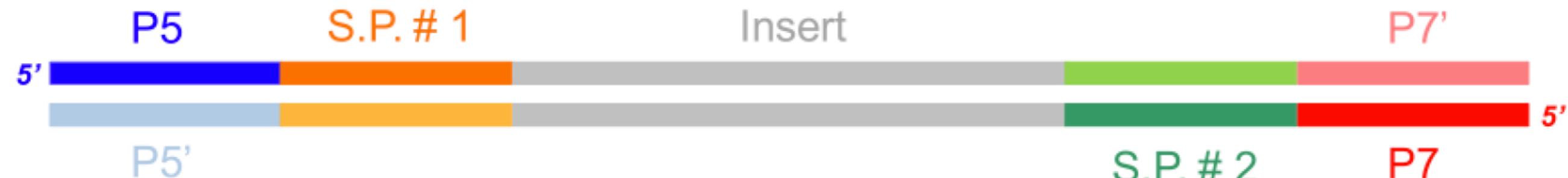


# ¿Qué es lo que necesitamos?

Single Read



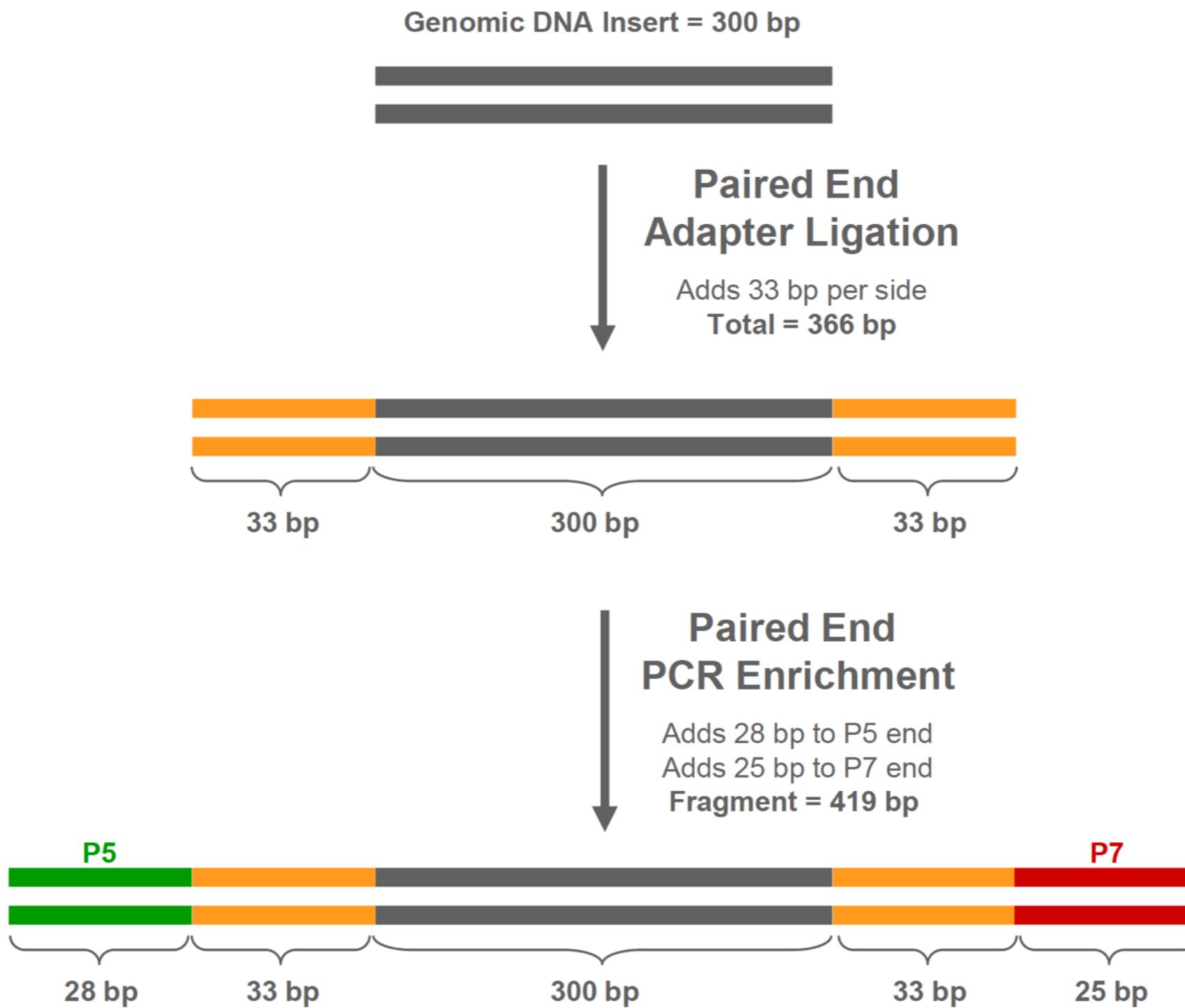
Paired Read



Indexed Paired Read

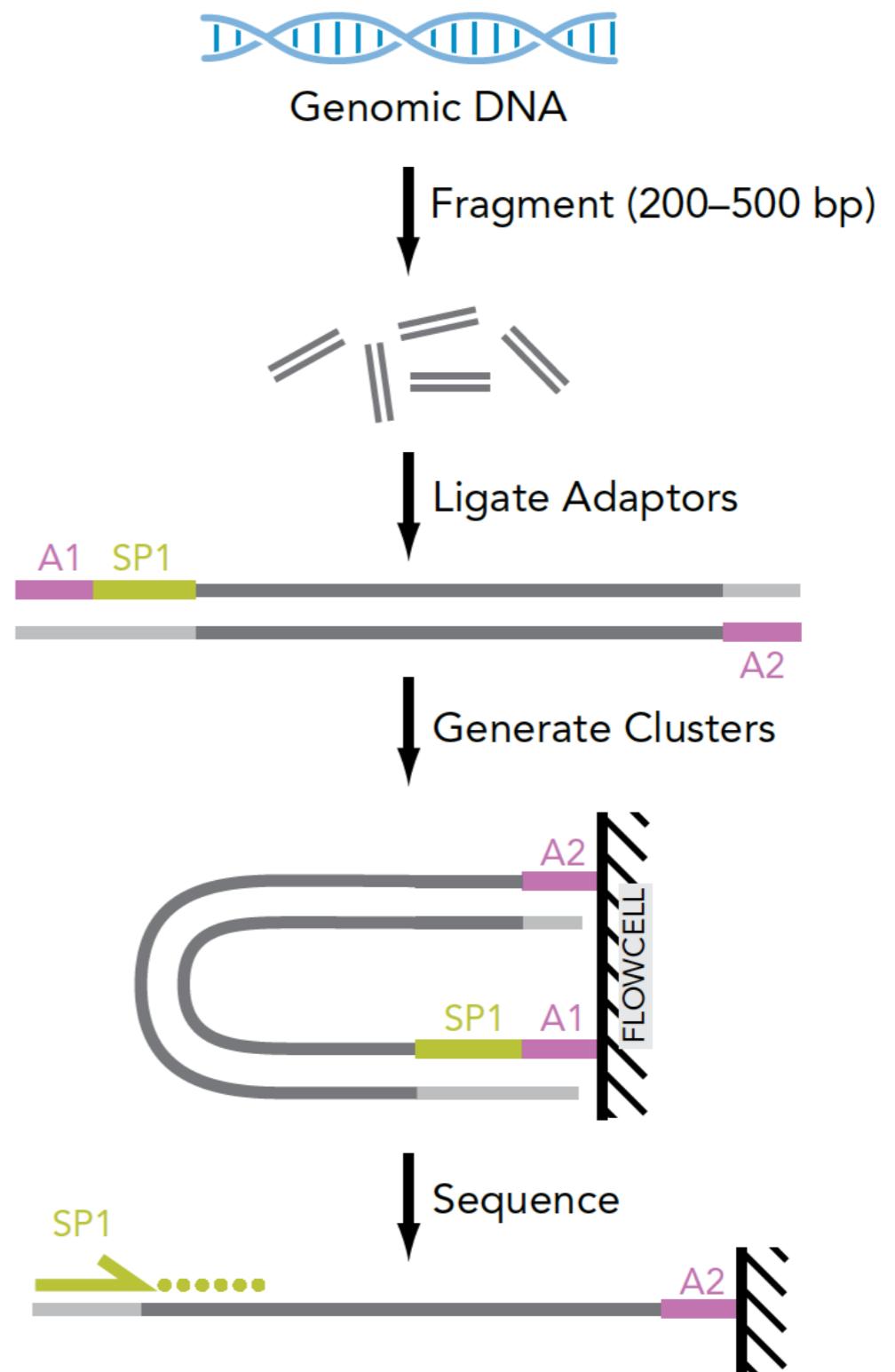


# Nota sobre tamaño de inserto y secuencia útil



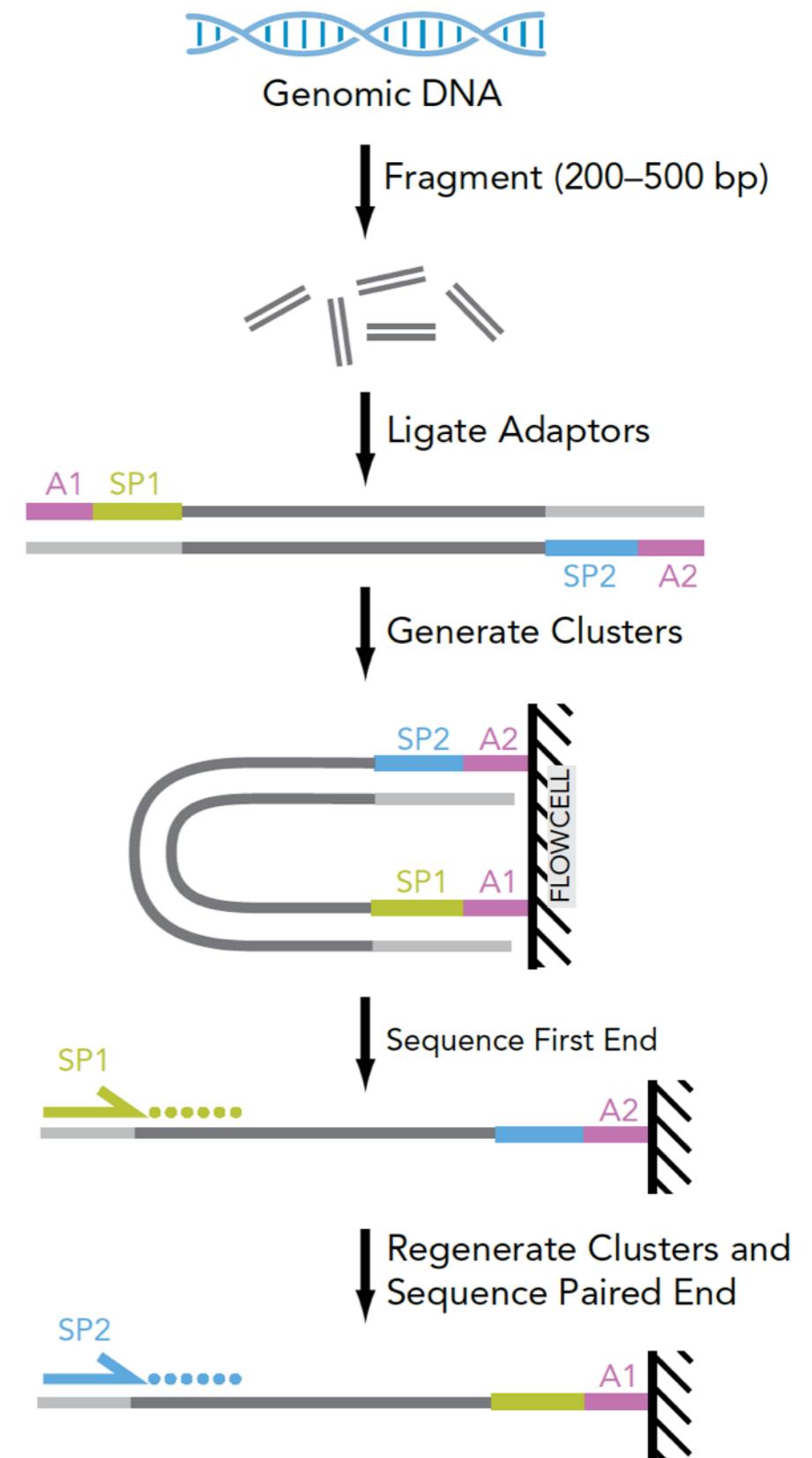
# Single-end

- Solo un partidor para secuenciar
- Rápido, más barato
- Descontinuado



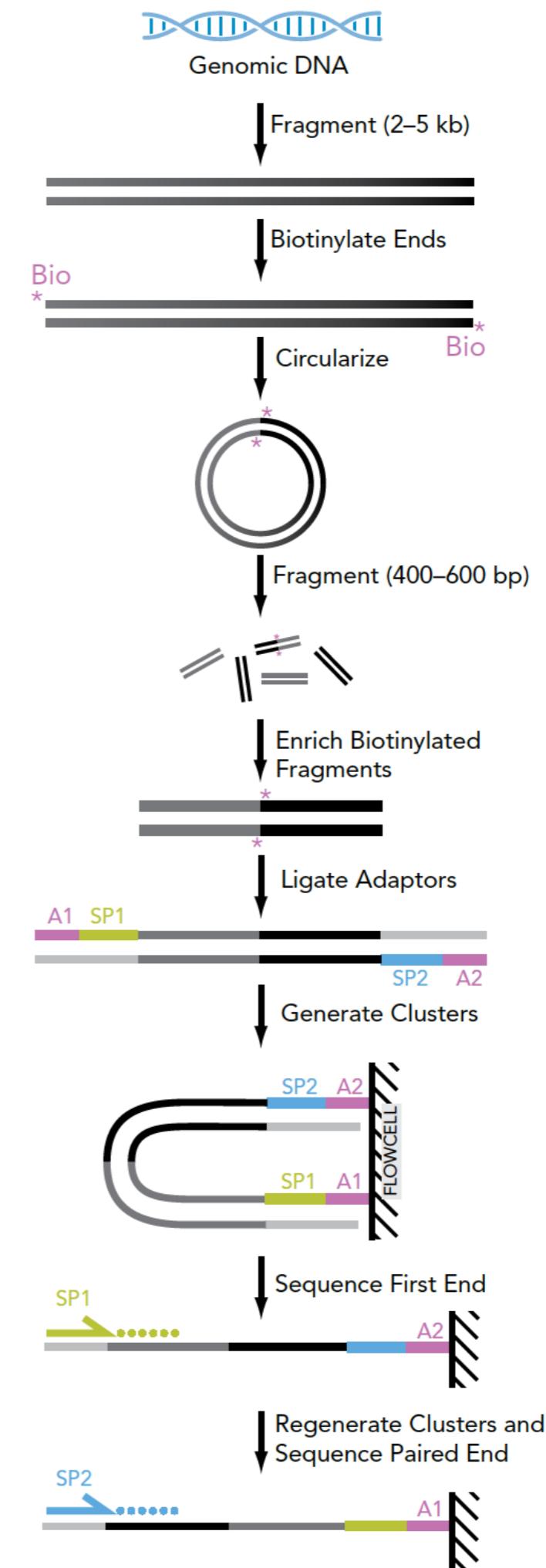
# Paired-end

- Se secuencia el mismo inserto dos veces
- Es posible “alargar” el tamaño de la read
- Captura información estructural
- Toma el doble de tiempo, más caro

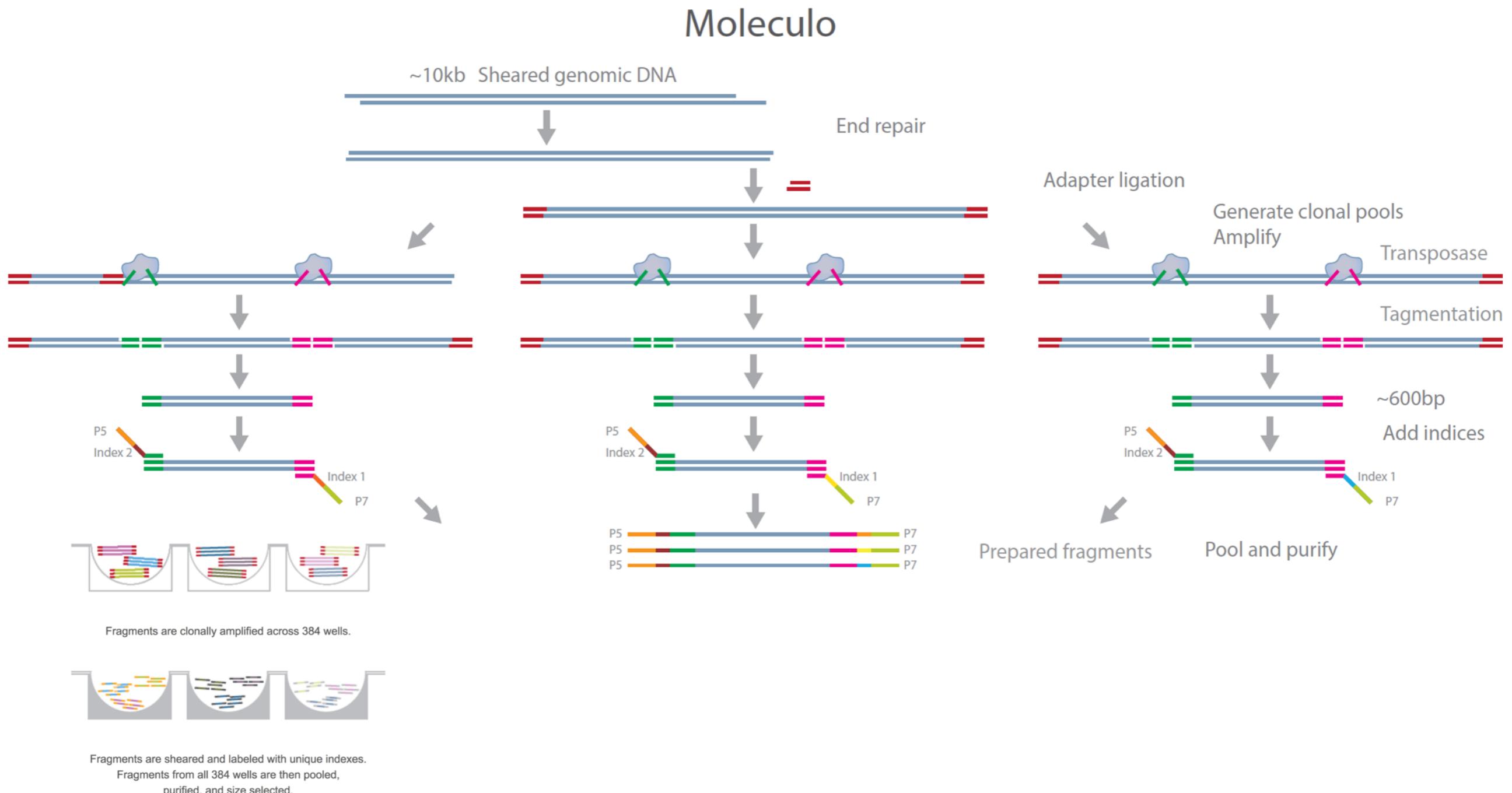


# Mate-pairs

- Información estructural
- Finalizar genomas, genomas de alta calidad
- Resolver genes multicopia, regiones repetitivas



# Artificial long reads



genomas complejos, “phasing” de alelos, finalizar genomas

# Después de la ligación viene la amplificación

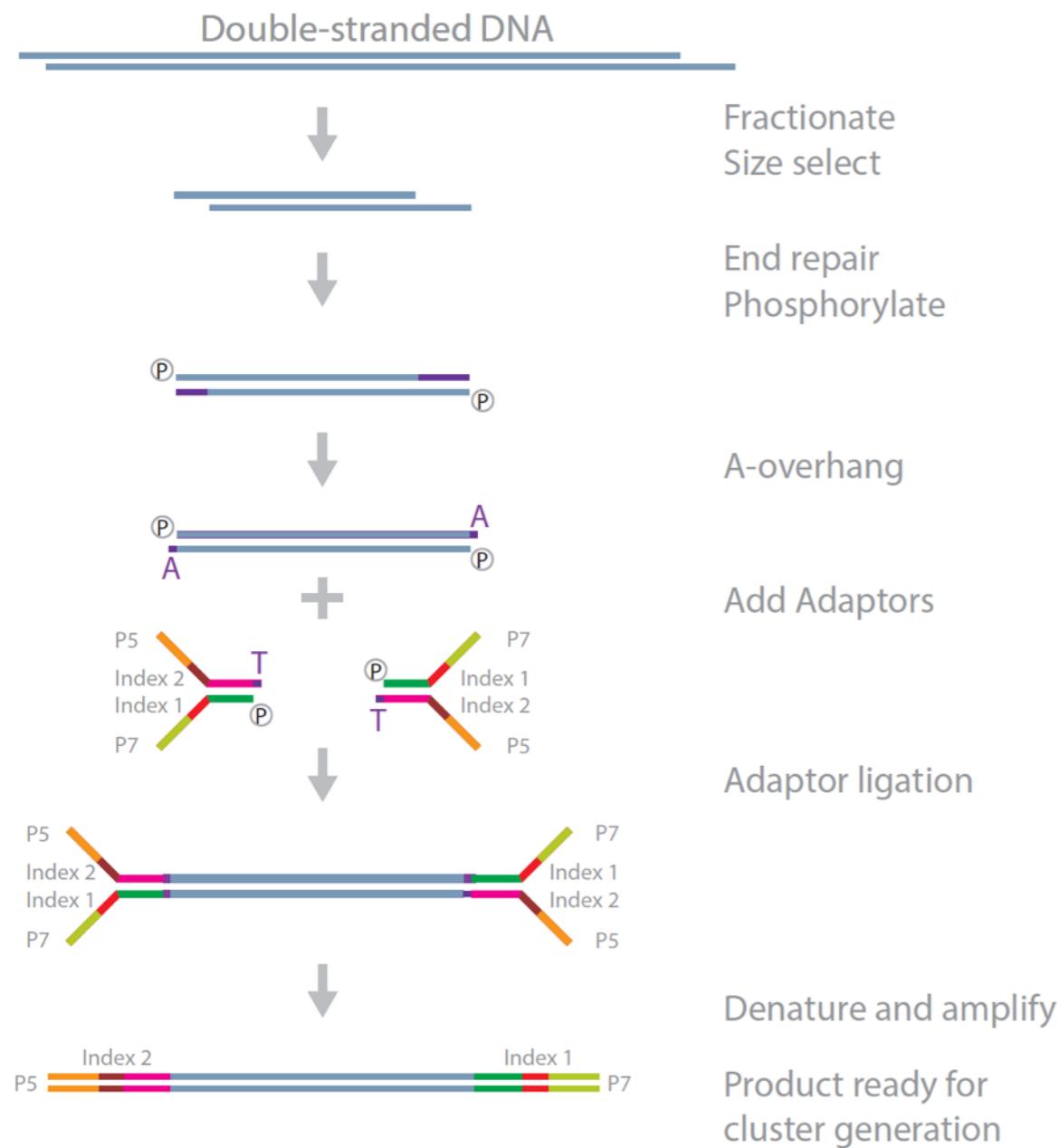
Selectively enrich DNA fragments with adapter molecules on both ends

Adds additional sequences to the end of the adapters for hybridization

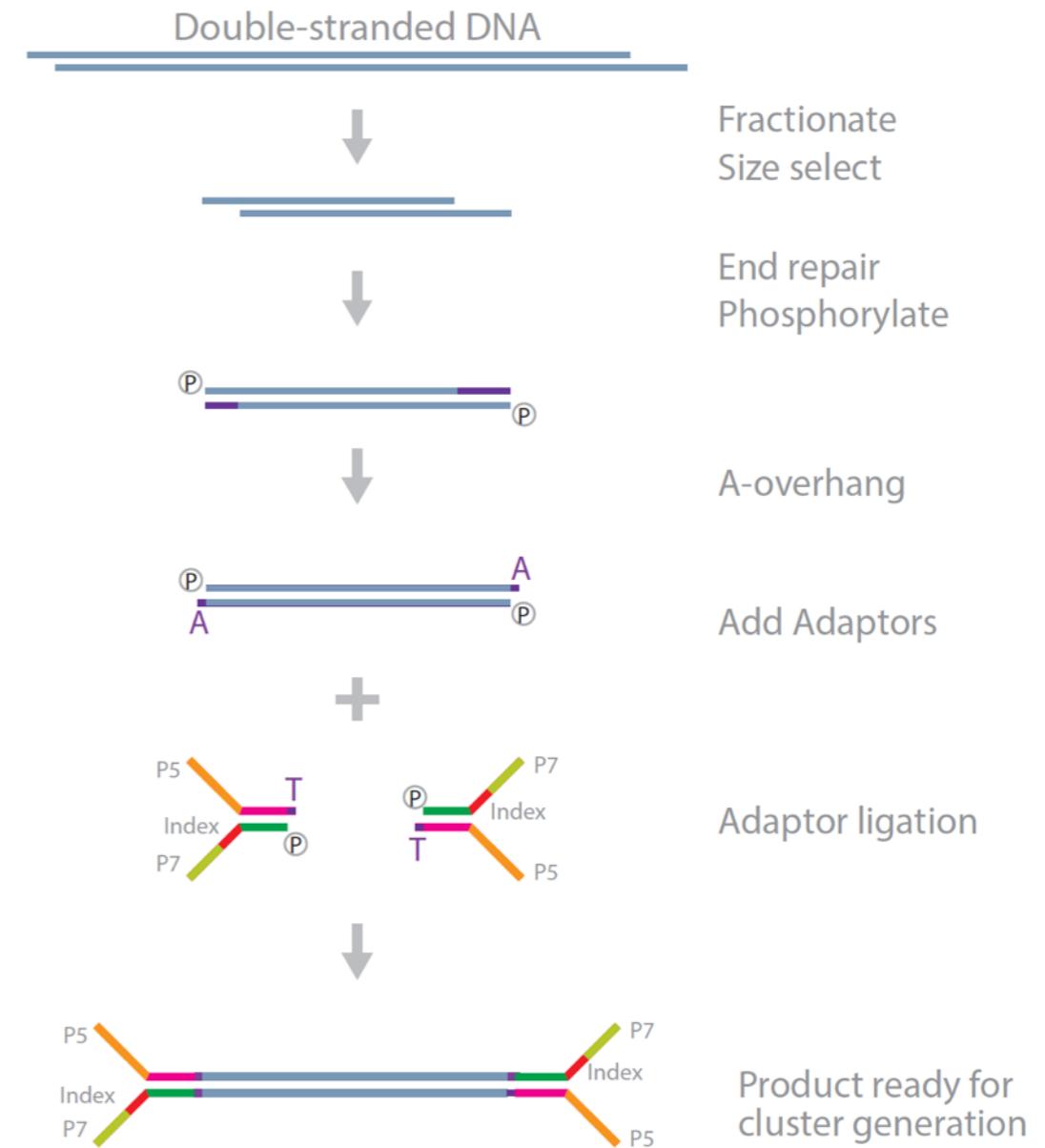
Amplifies the amount of DNA in the library

# Amplificar o no amplificar

## TruSeq Nano



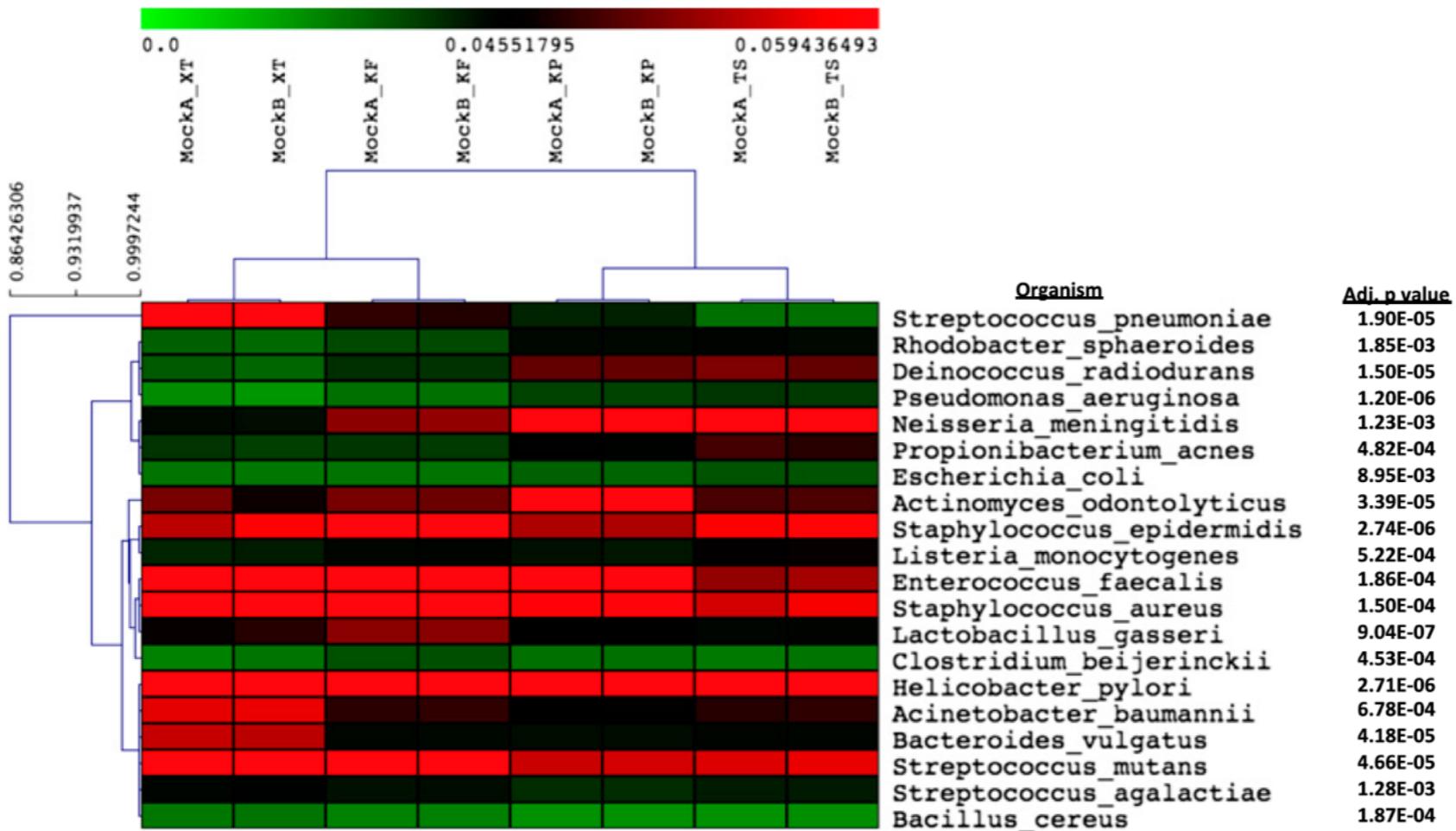
## TruSeq PCR Free



Menos input DNA, más sesgo

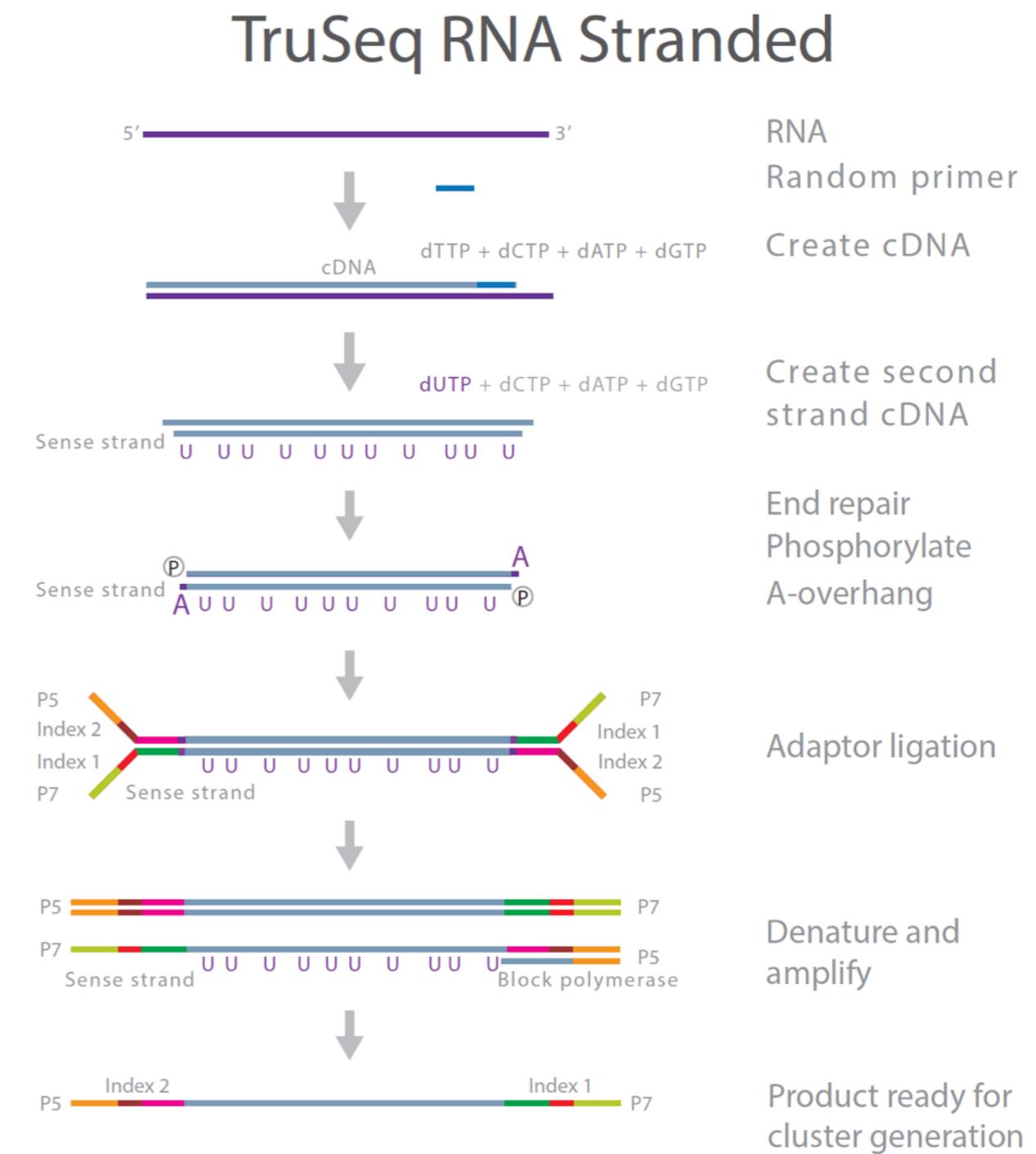
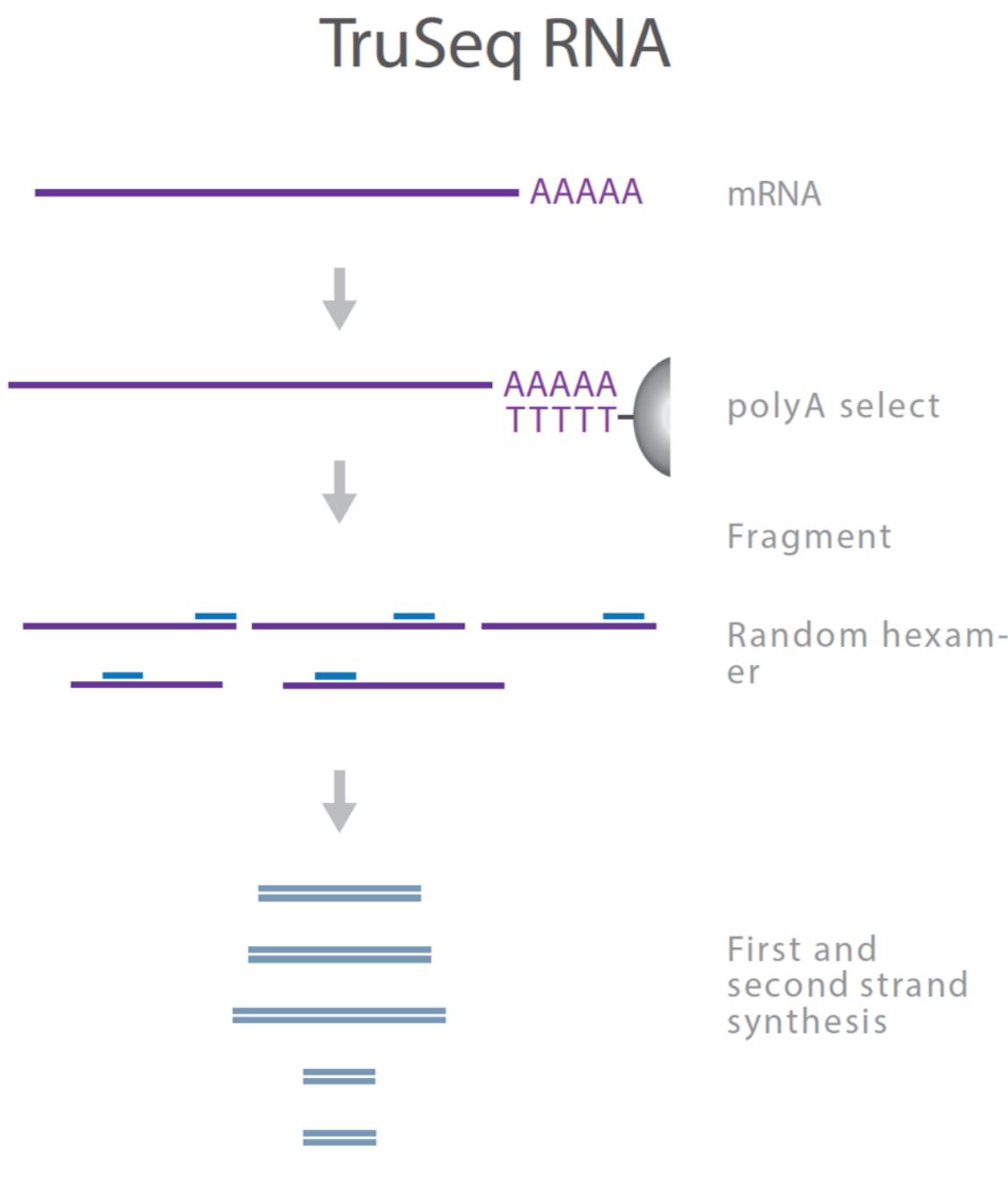
Más input DNA, menos sesgo

# Amplificar o no amplificar



**Fig. 1.** One-way ANOVA analysis across library preparation methods. Relative abundance measurements were calculated for the mock community across the four different protocols and analyzed for consistency between library preparations from both technical replicates. Shading in the heat map indicates relative abundance in the mock-community DNA mixture from low (green) to high (red) abundance. Adjusted *P* values were calculated based on a maximum *P* value of 0.01. Samples and organisms were clustered based on an uncentered Pearson complete linkage analysis. The letters "A" and "B" indicate technical replicates for each sample preparation.

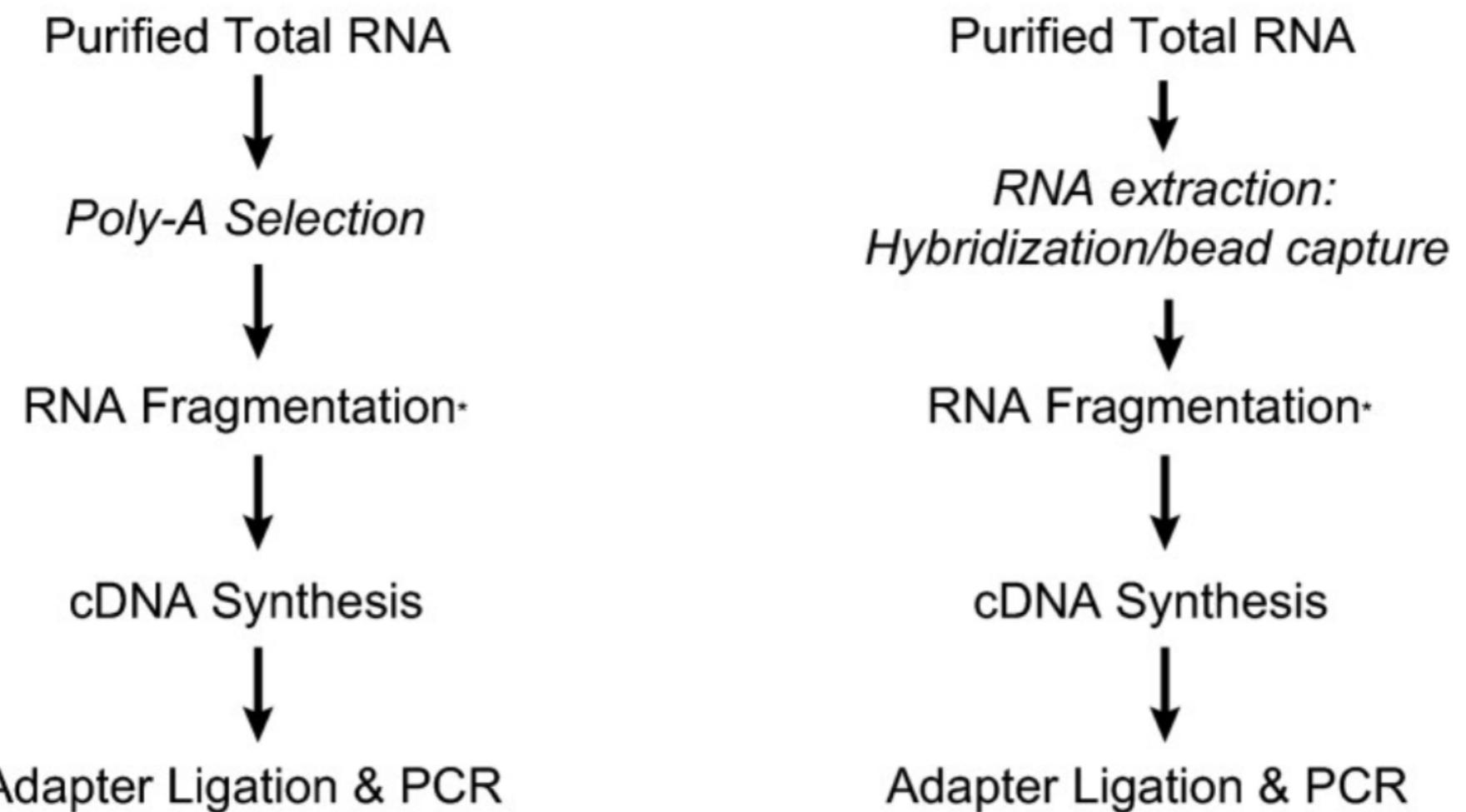
# RNASeq: Stranded o regular



¿De cuál hebra el RNA está siendo transcrita?

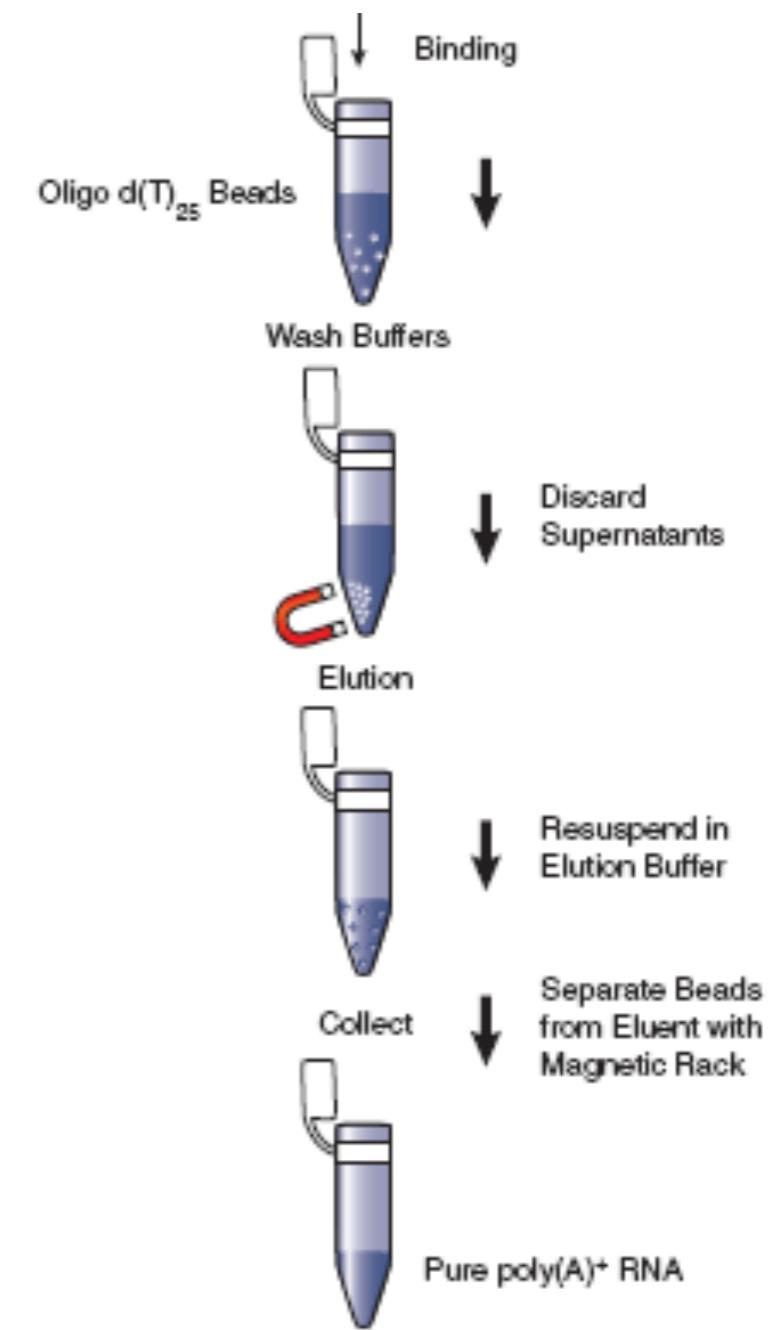
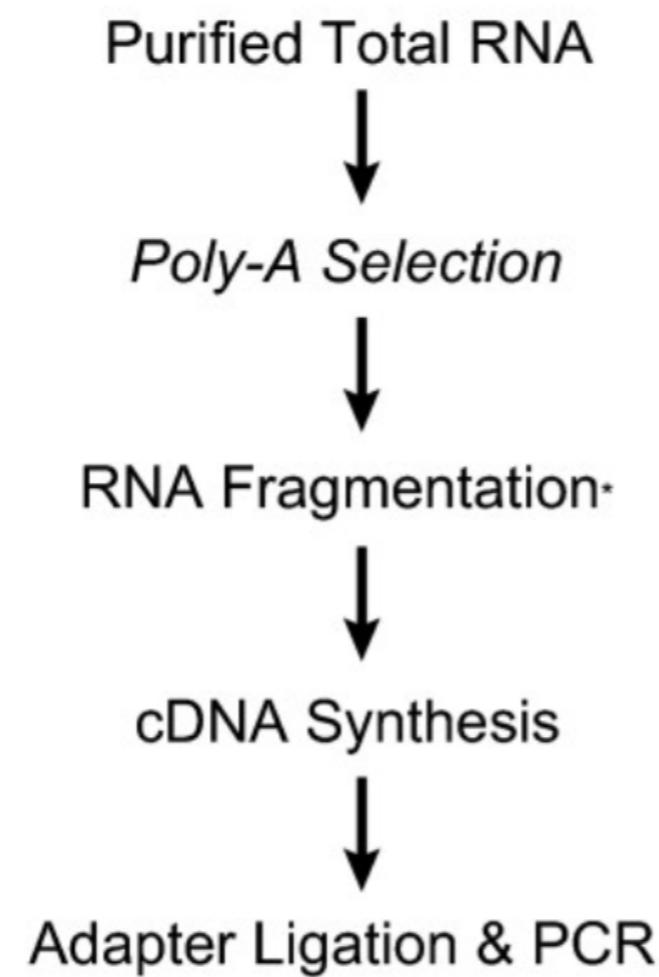
# RNASeq: ¿Cómo seleccionar mensajeros?

- Poly-A
- Remover rRNAs



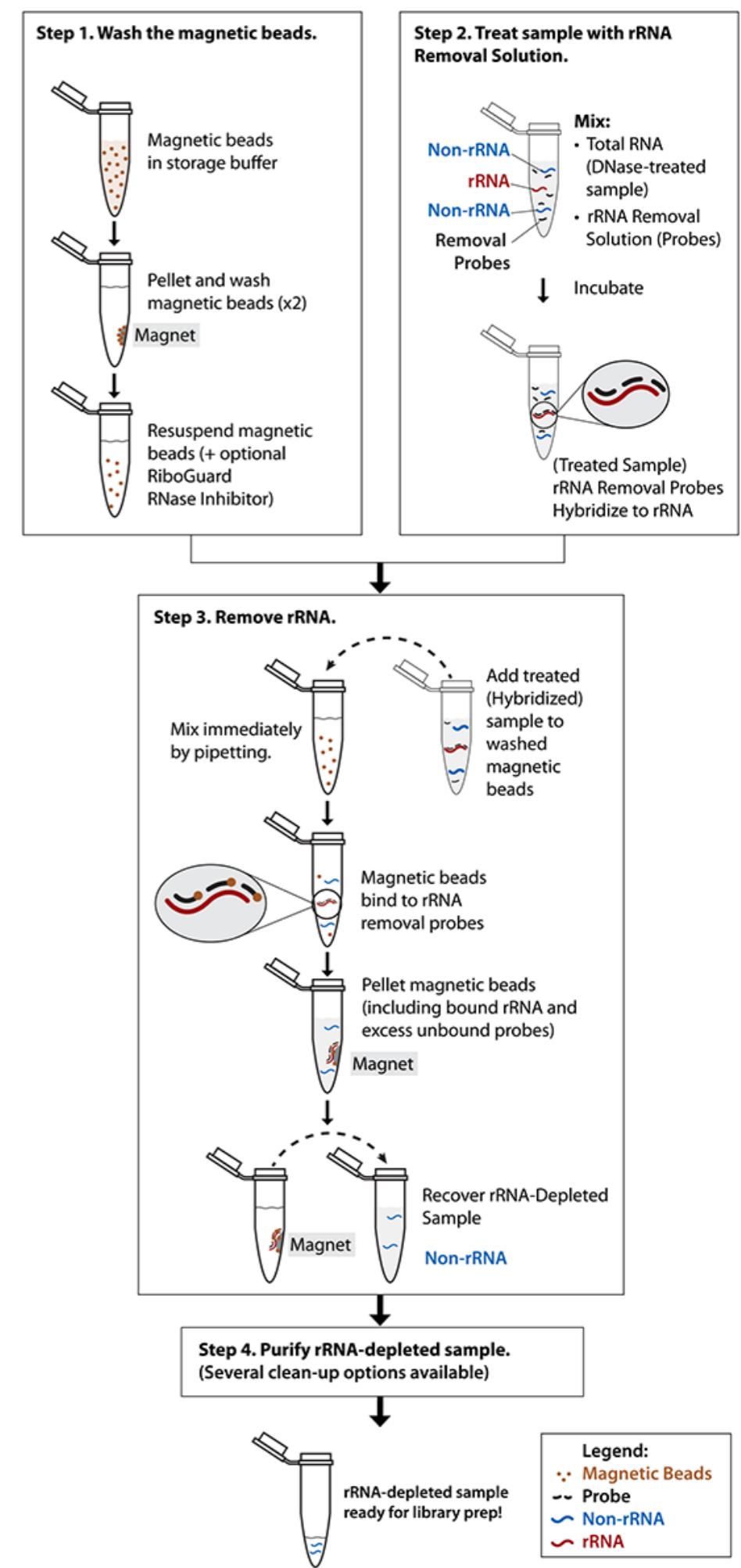
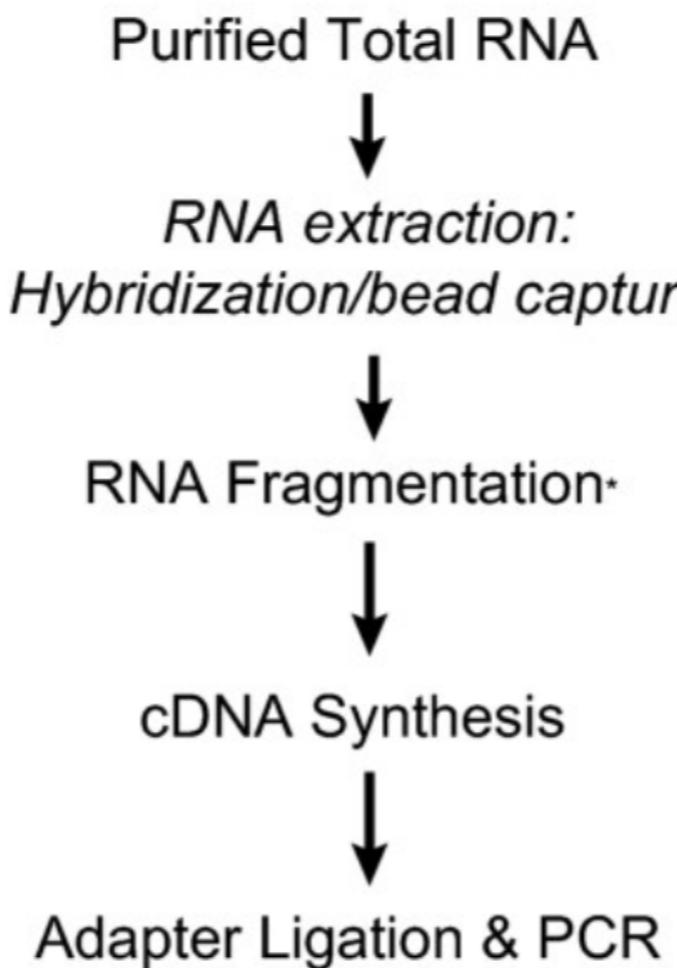
# Selección por Poly A

- Poly-A
- Rápido, barato
- Solo mRNA, no lncRNA u otras especies



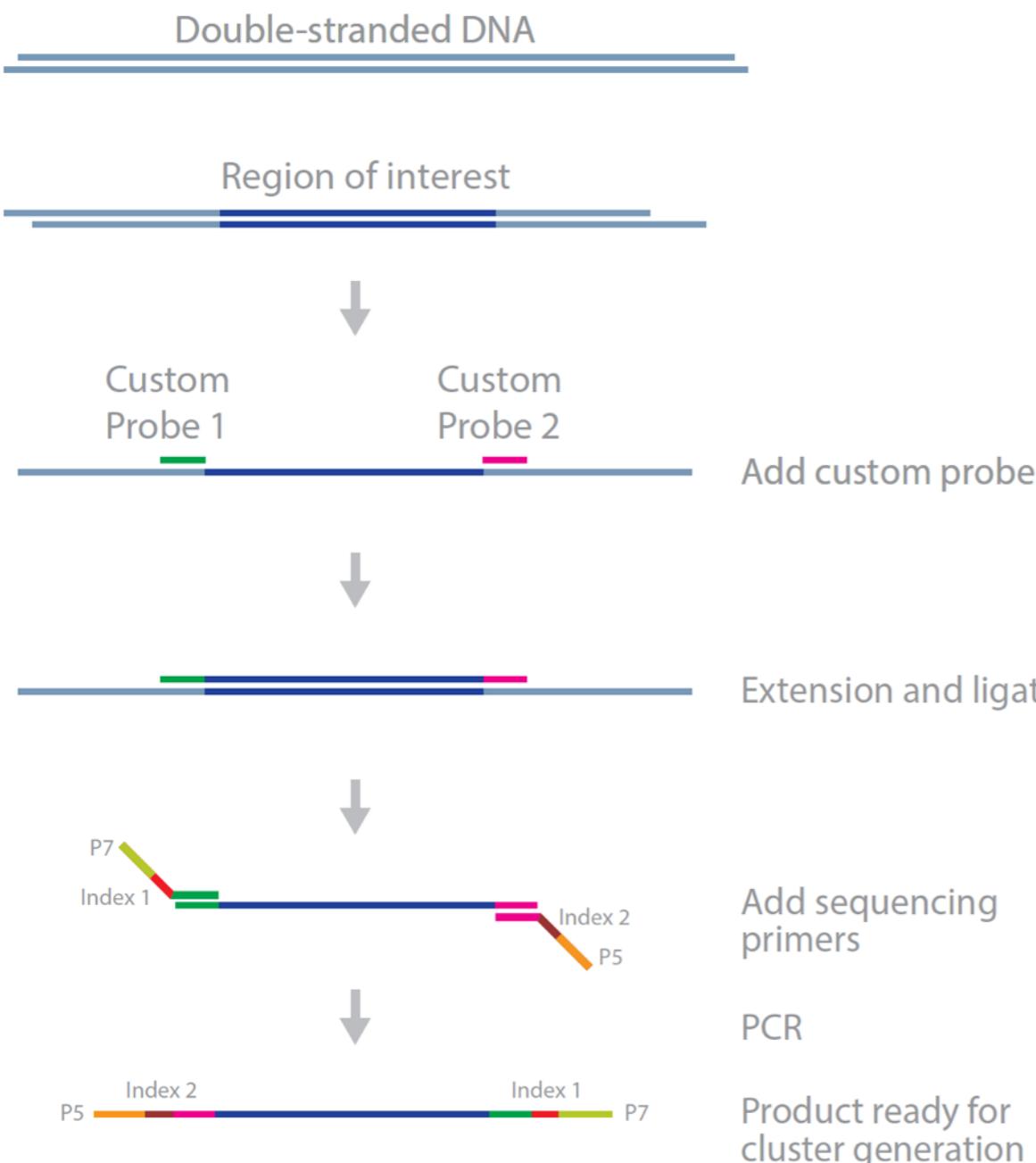
# Selección por remoción de rRNA

- Sondas contra rRNA
- Rápido, caro
- mRNA + lncRNA y otras especies



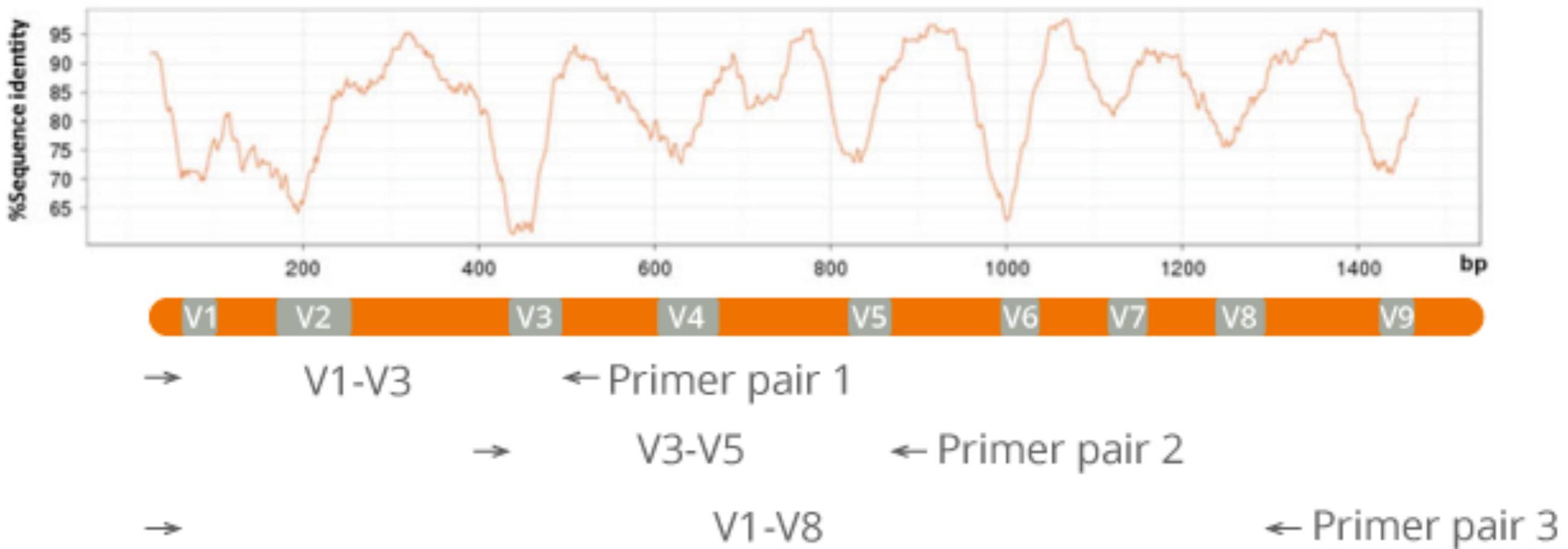
# Secuenciar productos de PCR

## TruSeq Custom Amplicon

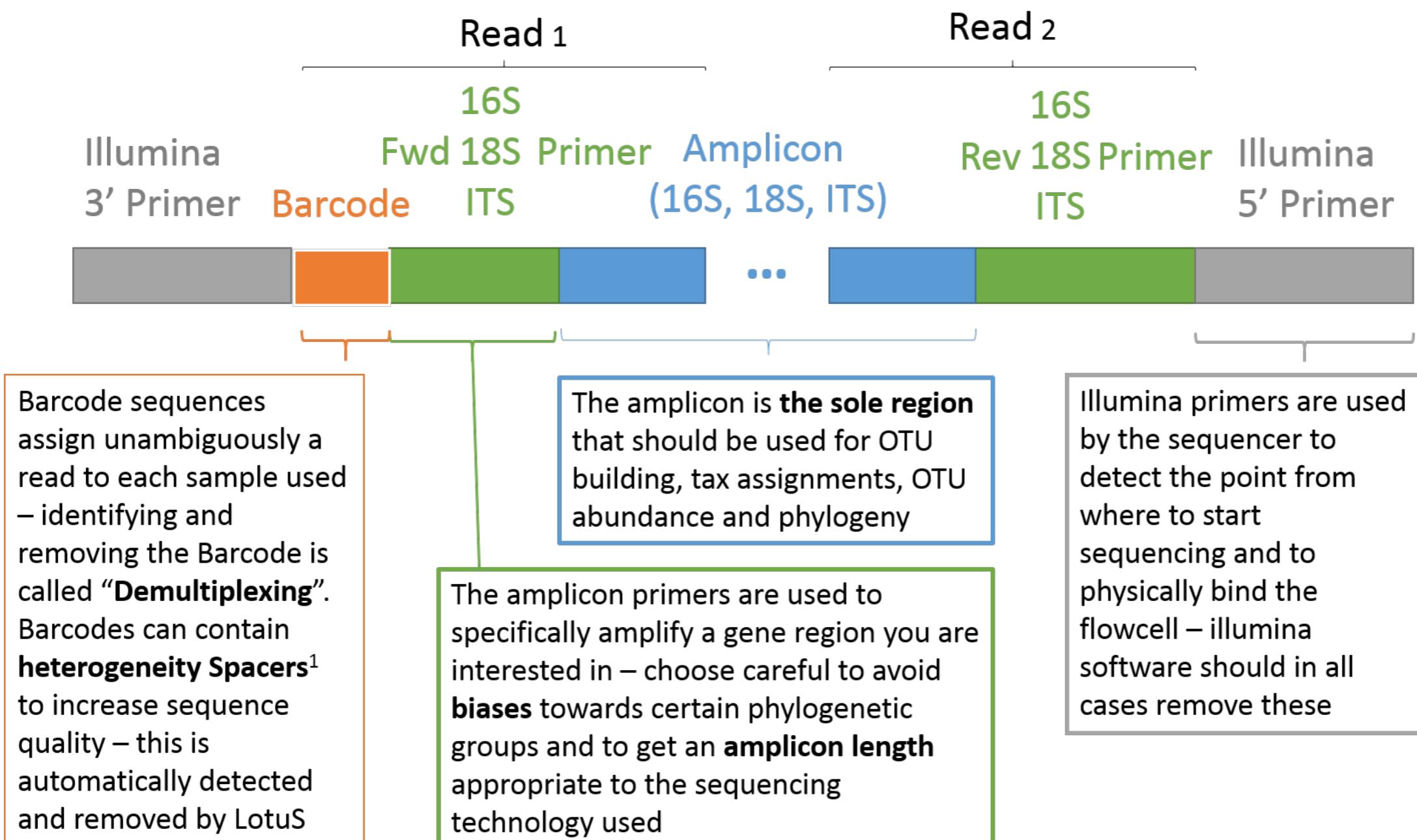


- Genes de interés
- Genes candidato para enfermedades
- Marcadores taxonómicos, 16S, 18S, ITS

# Secuenciar productos de PCR: 16S rRNA gene

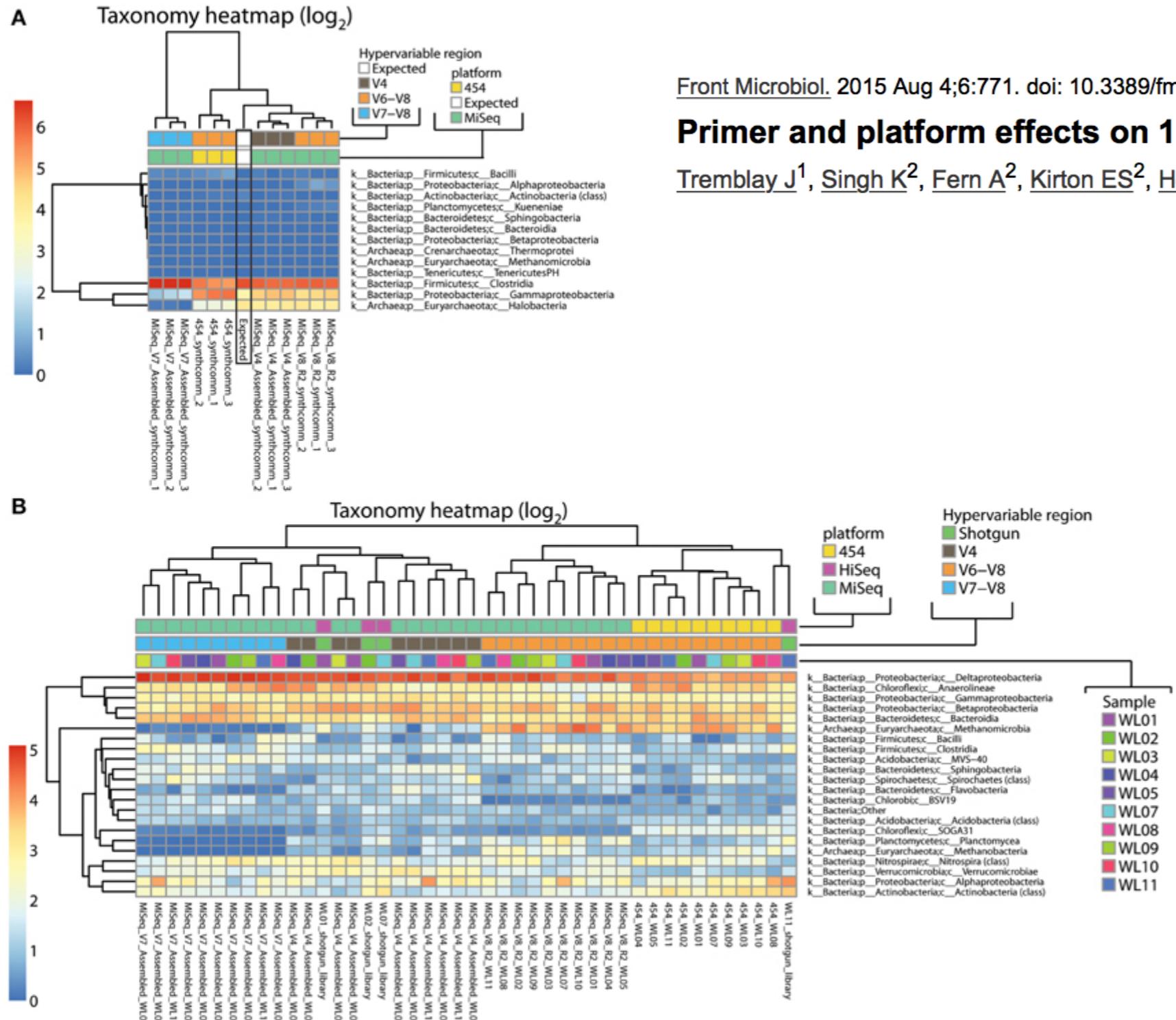


# Secuenciar productos de PCR: 16S rRNA gene



1) Fadrosh DW, Ma B, Gajer P, Sengamalay N, Ott S, Brotman RM, Ravel J. 2014. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome* 2: 6.

# Secuenciar productos de PCR: 16S rRNA gene



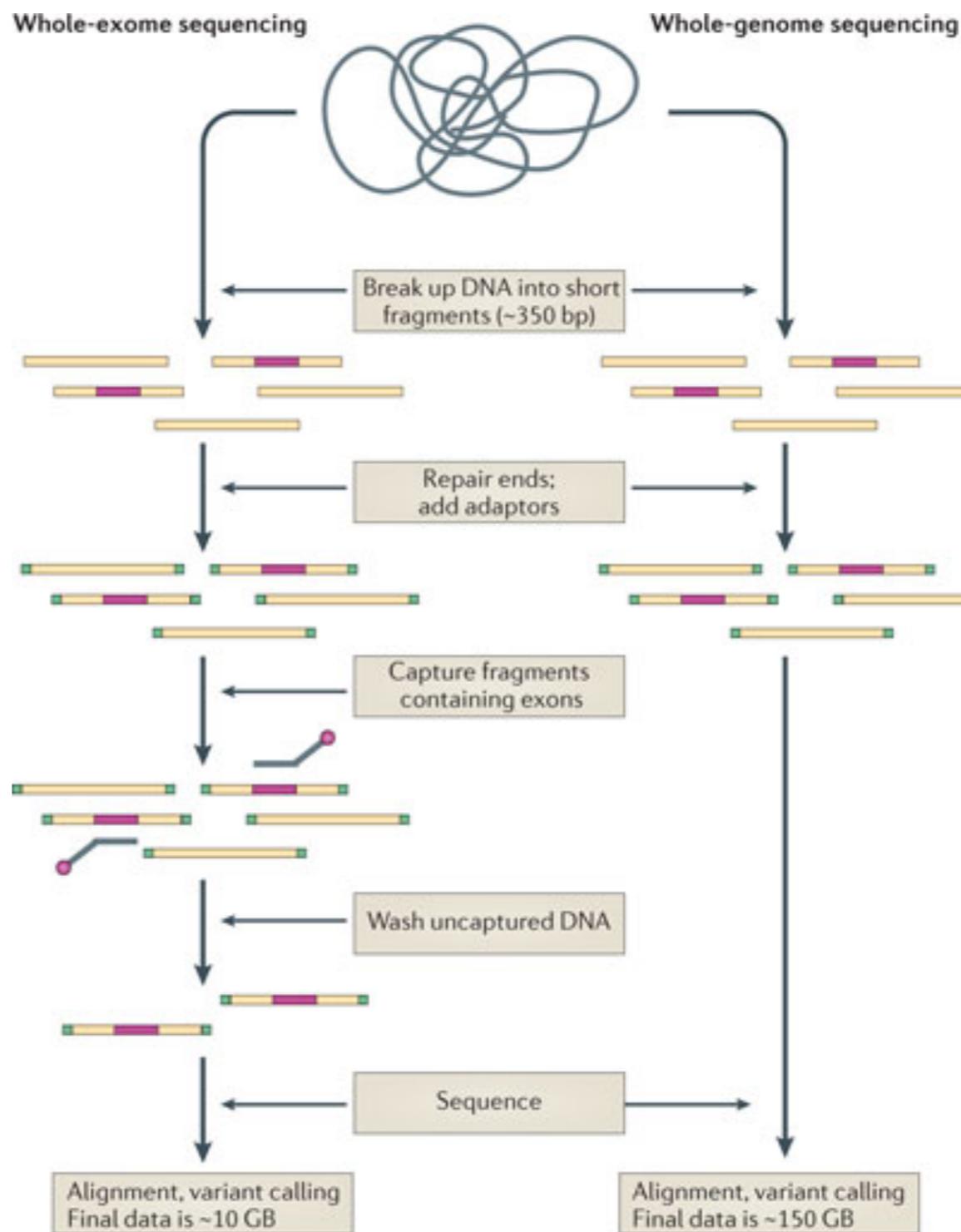
# Estrategias especiales

# Estrategías de representación reducida o de captura

- Exones
- Elementos Ultra-Conservados
- VirusCap
- ChIPSeq
- Captura por sondas conjugadas
  - Sondas son específicas para la población target

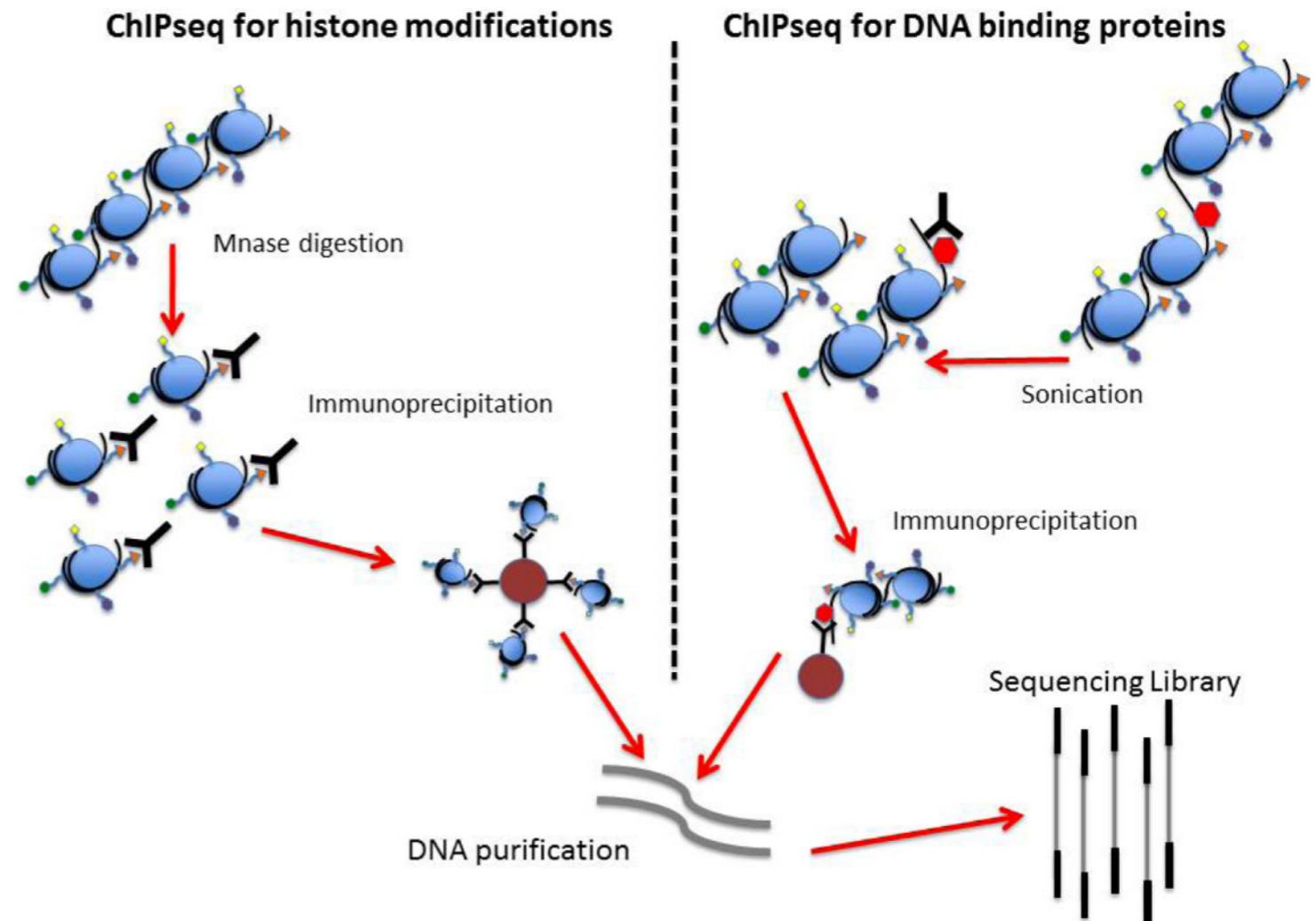
# Estrategías de representación reducida o de captura

- Exones



# Estrategías de representación reducida o de captura

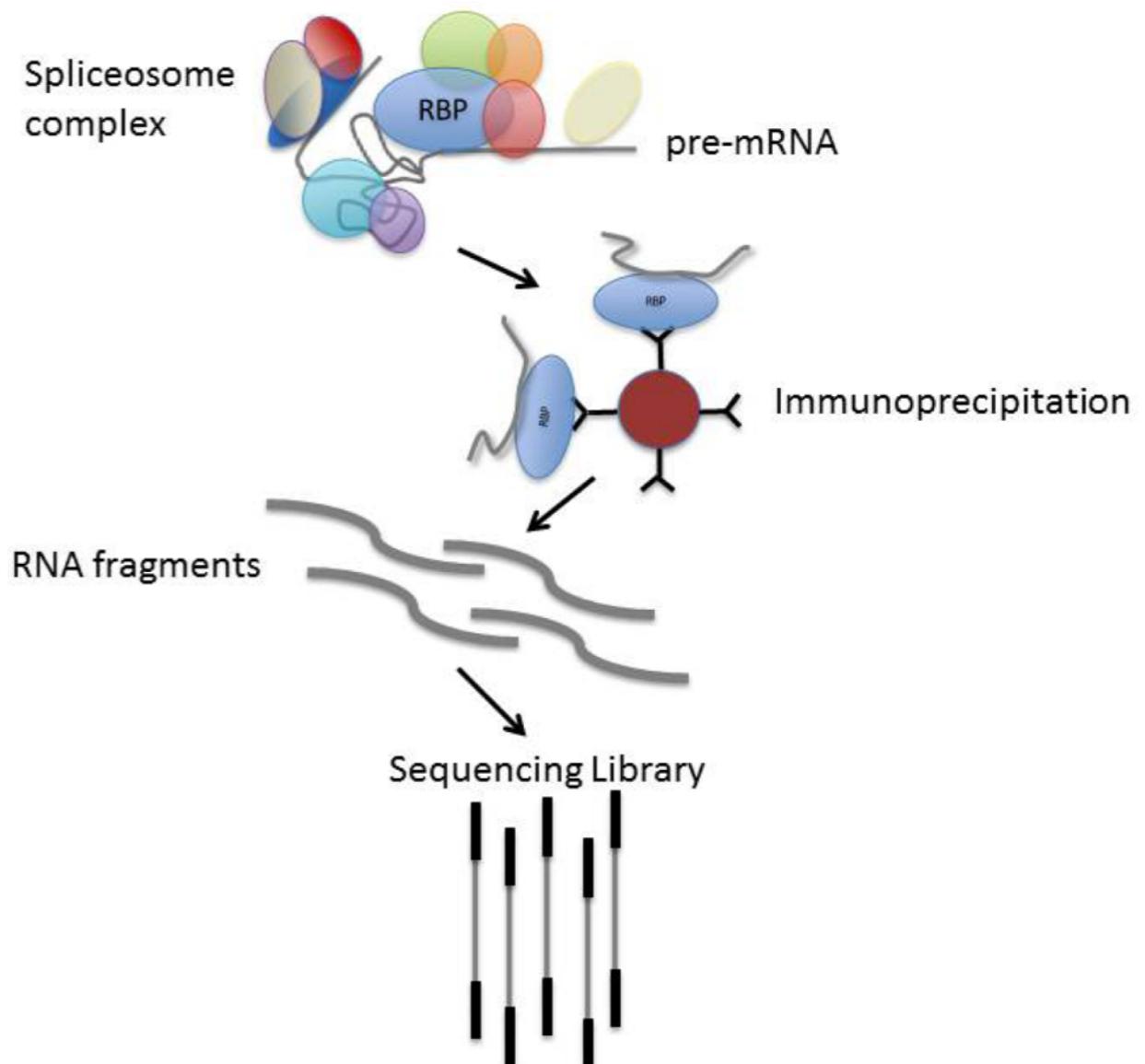
- Estructura de la cromatina
- Spliceosome



**Figure 5. ChIP-seq procedure for detecting sequences at the sites of histone modifications or the recognition sequences of DNA binding proteins.** Chromatin is crosslinked, fragmented either by micrococcal nuclease digestion or by sonication, and then incubated with antibodies for either the histone modification or protein of interest. Immunoprecipitation is performed using either Protein A or Protein G beads. After washing, the DNA is uncrosslinked, eluted from the beads and purified, at which point the DNA can be taken into standard DNA library construction protocols.

# Estrategías de representación reducida o de captura

- Estructura de la cromatina
- Spliceosome
- Basados en cross-linking reversible + inmunoprecipitación



**Figure 6. RNA immunoprecipitation (RIP-seq) done by targeting RNA binding proteins (RBPs).** The basic principle of RIP-seq is immunoprecipitation of RBPs that are bound to target RNA molecules. The RNA molecules are then purified and a sequencing library is created. In some protocols, the RBP complex is chemically crosslinked to the target RNA; that crosslinking must be reversed after immunoprecipitation. We have found that crosslinking is not necessary for simple RIP-seq where the objective is to identify the RNA molecules bound by RBP, but it is required for CLIP-seq protocols that are used to identify the specific sequence motifs for RBP binding. The immunoprecipitation step can be done with antibodies directed at the specific RBP of interest, or the RBP can be tagged and expressed in the cells under study.

# Control de calidad

# Control de calidad de librerías

- Complejidad de la muestra - qué tan diversos o representativos son los insertos en la librería
- Afectan la complejidad: cantidad de material de partida, cantidad de DNA perdido durante limpieza y selección, cantidad de duplicación introducida por PCR

# Control de calidad de librerías

- Asegurarse de que la distribución de tamaños sea apropiada
- Al menos 20 nM de librería

# Control de calidad de librerías: después de secuenciar

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.