NuGEN
imagine more from less®

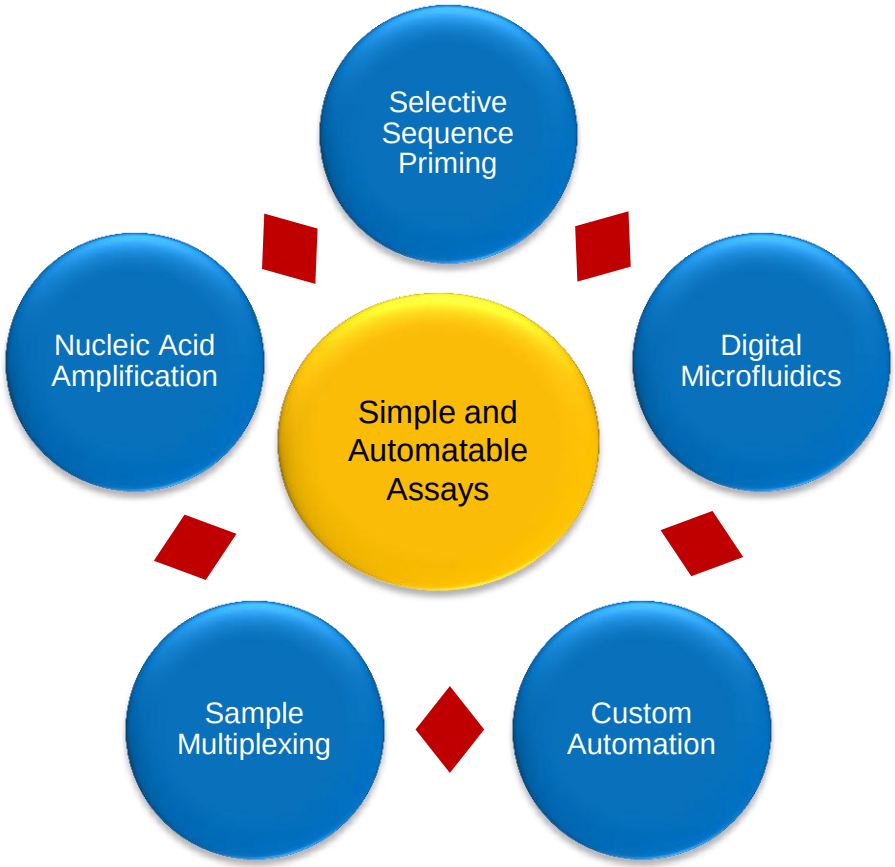
Integrated NGS Sample Preparation Solutions for Limiting Amounts of RNA and DNA

March 2, 2013

Steven R. Kain, Ph.D.

ABRF 2013

NuGEN's Core Technologies



NuGEN
imagine more from less®

The Mondrian™ SP+ System

Using digital microfluidics for simple, reliable and cost effective automation of genomic sample preparation.

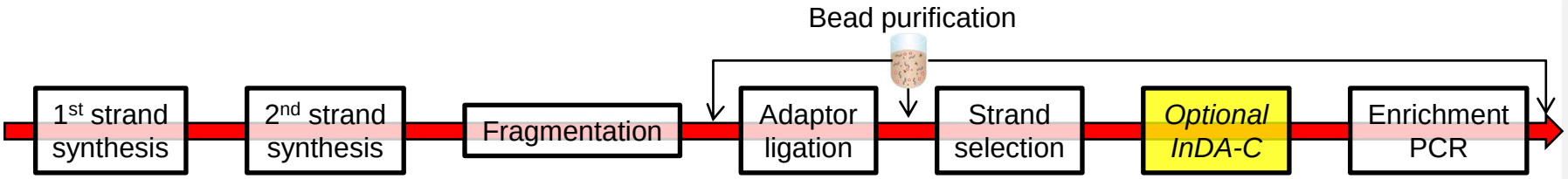
NuGEN
imagine more from less®

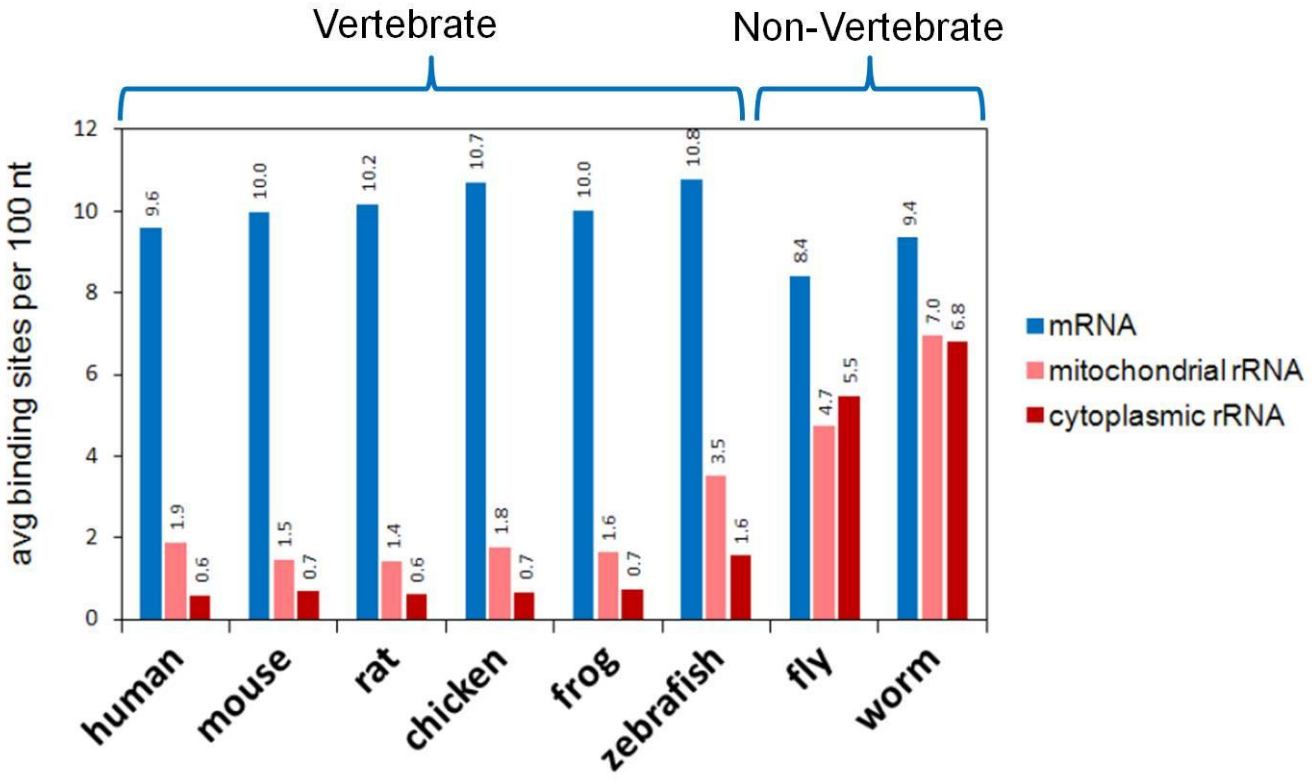
Encore Complete RNA-Seq Library Systems

- **Complete solutions for strand-specific RNA-Seq** – All required components for preparation of whole transcriptome *strand-specific* RNA-Seq libraries
- **Simple, fast, automatable workflow** – Library construction starting with total RNA in as little as seven hours, with only three purification steps
- **Affordable and scalable** – Optional barcoding for multiplex sequencing on the Illumina platforms

NuGEN
imagine more from less®

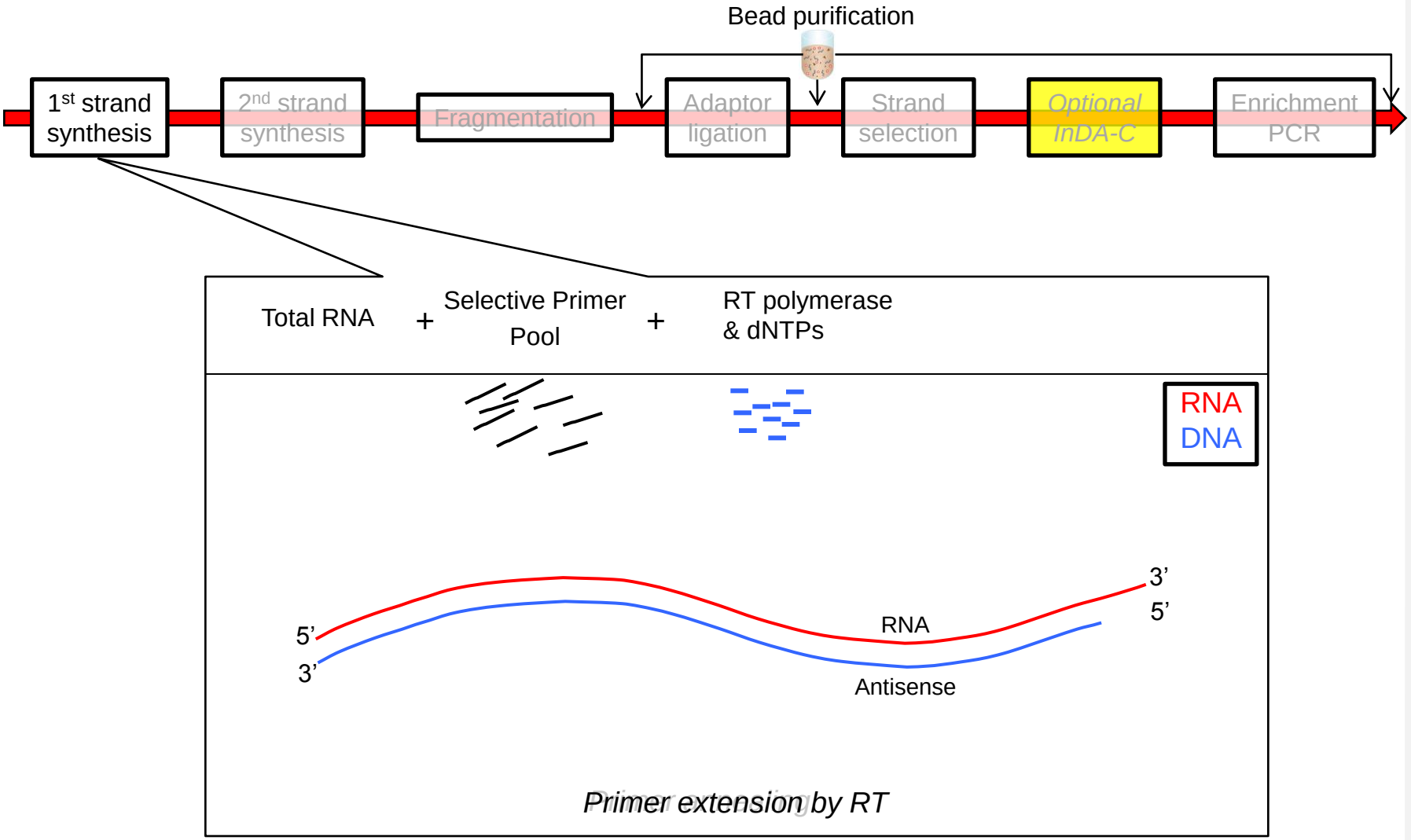
Strand-Specific RNA-Seq





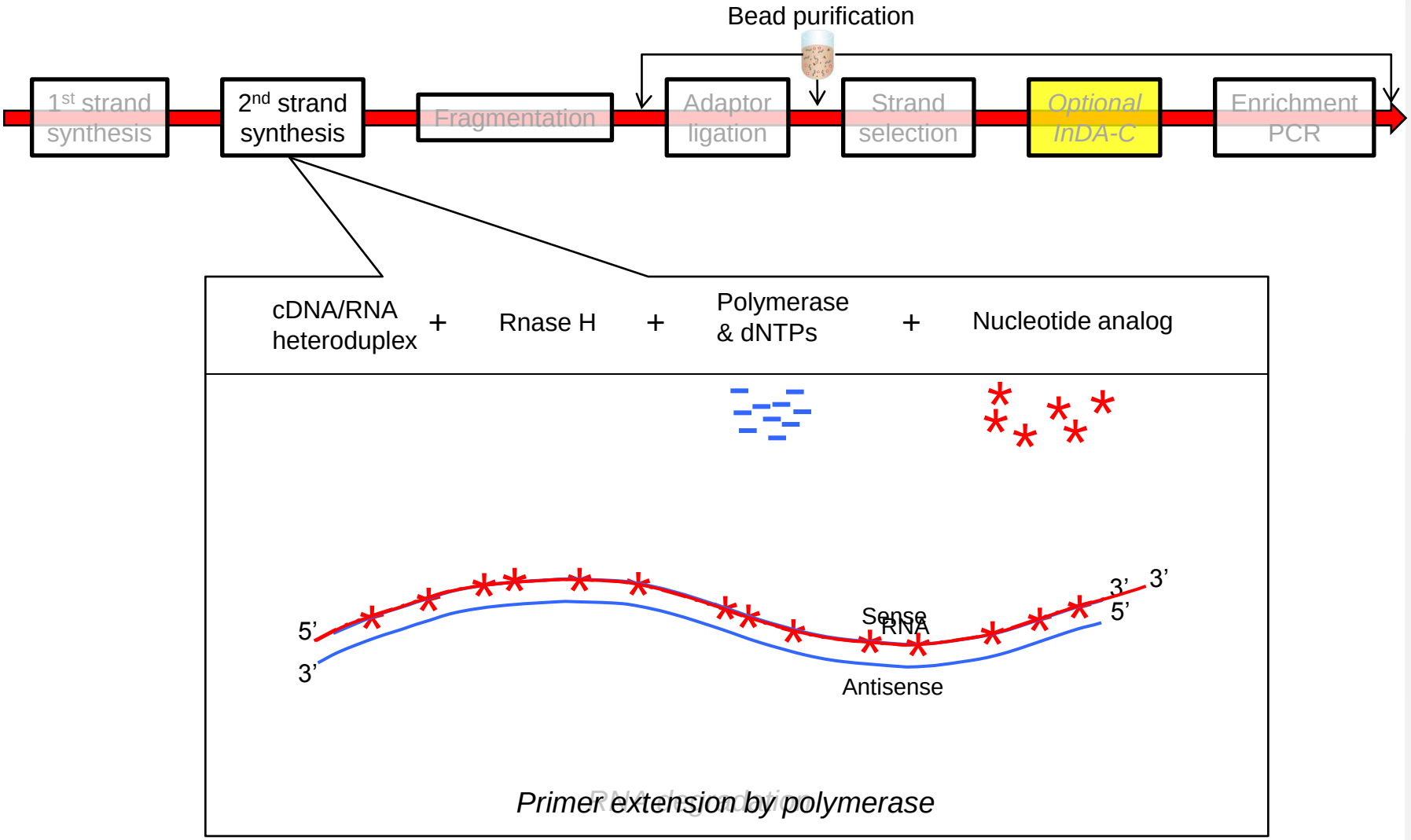
NuGEN imagine more from less®

Strand-Specific RNA-Seq: Selective Priming



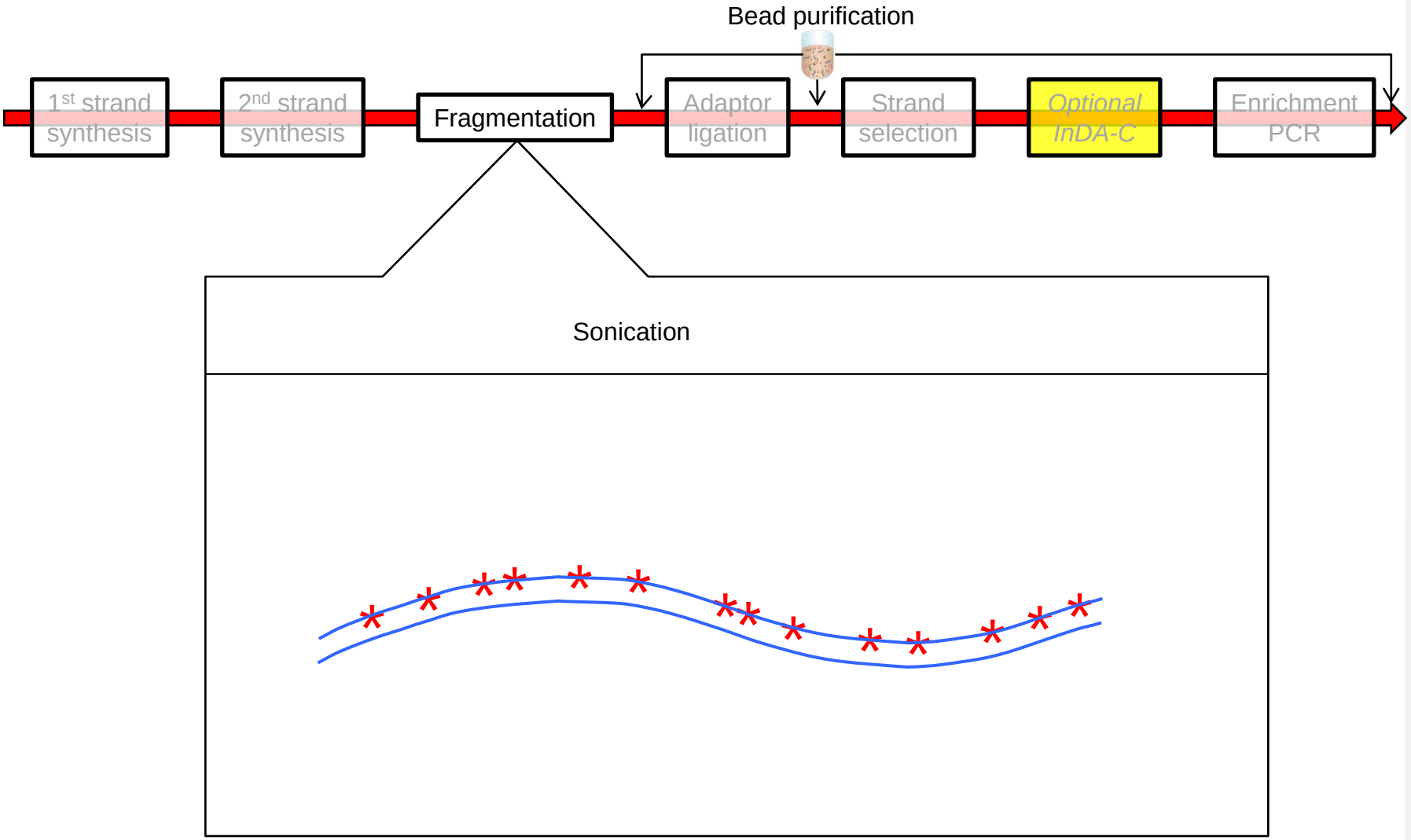
NuGEN imagine more from less®

Strand-Specific RNA-Seq: Sense-Strand Tagging

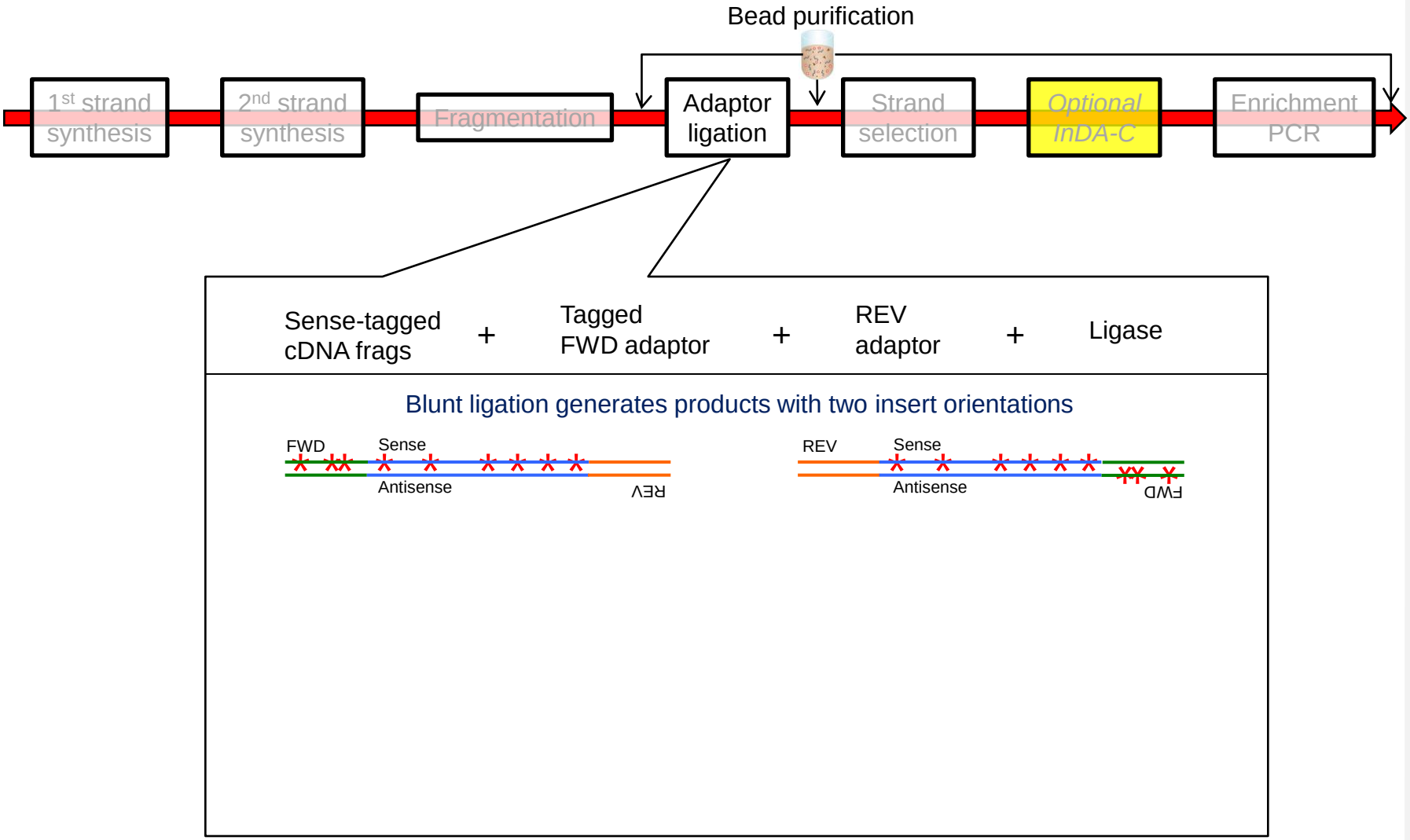


NuGEN imagine more from less

Strand-Specific RNA-Seq: cDNA Fragmentation

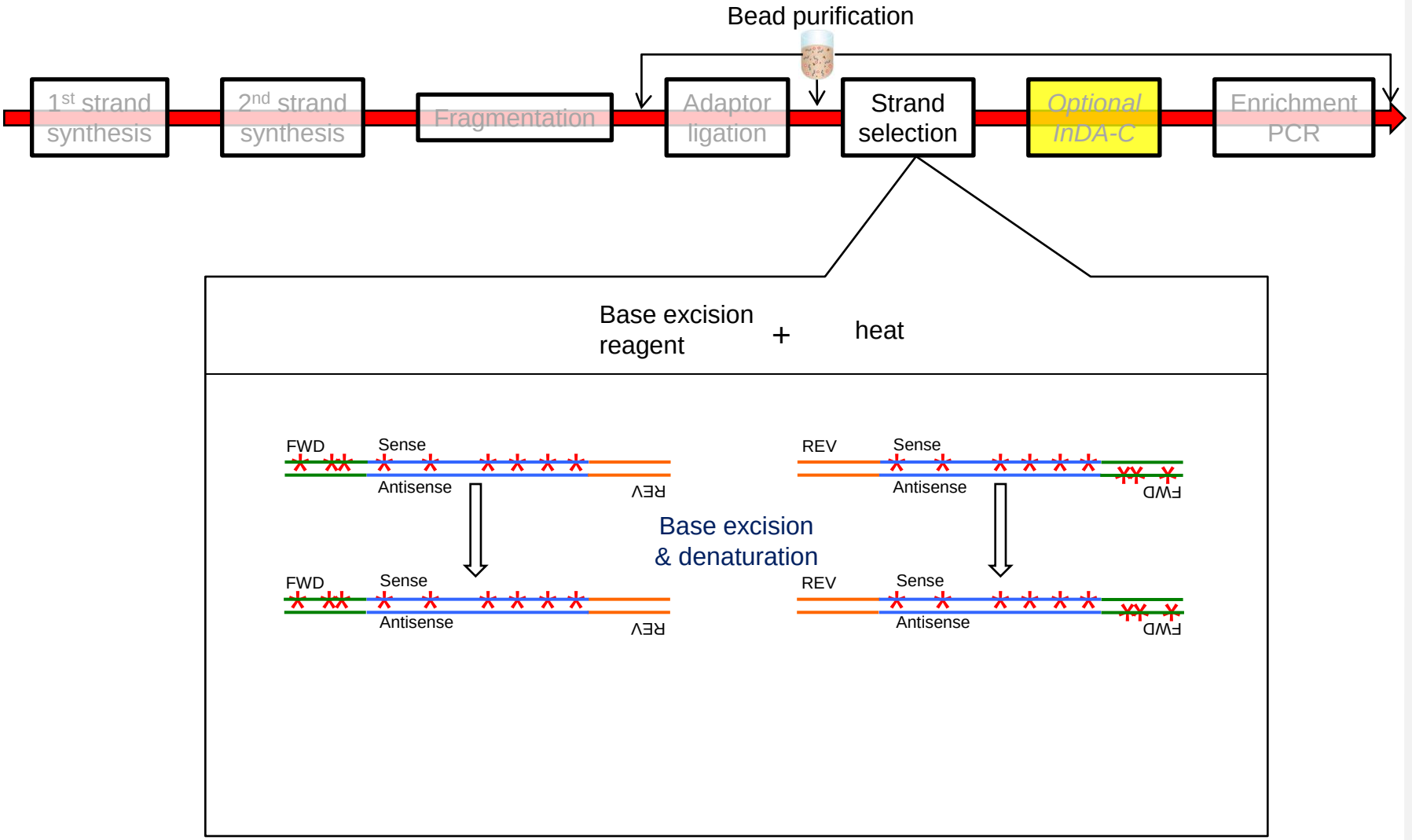


Strand-Specific RNA-Seq: Adaptor Ligation



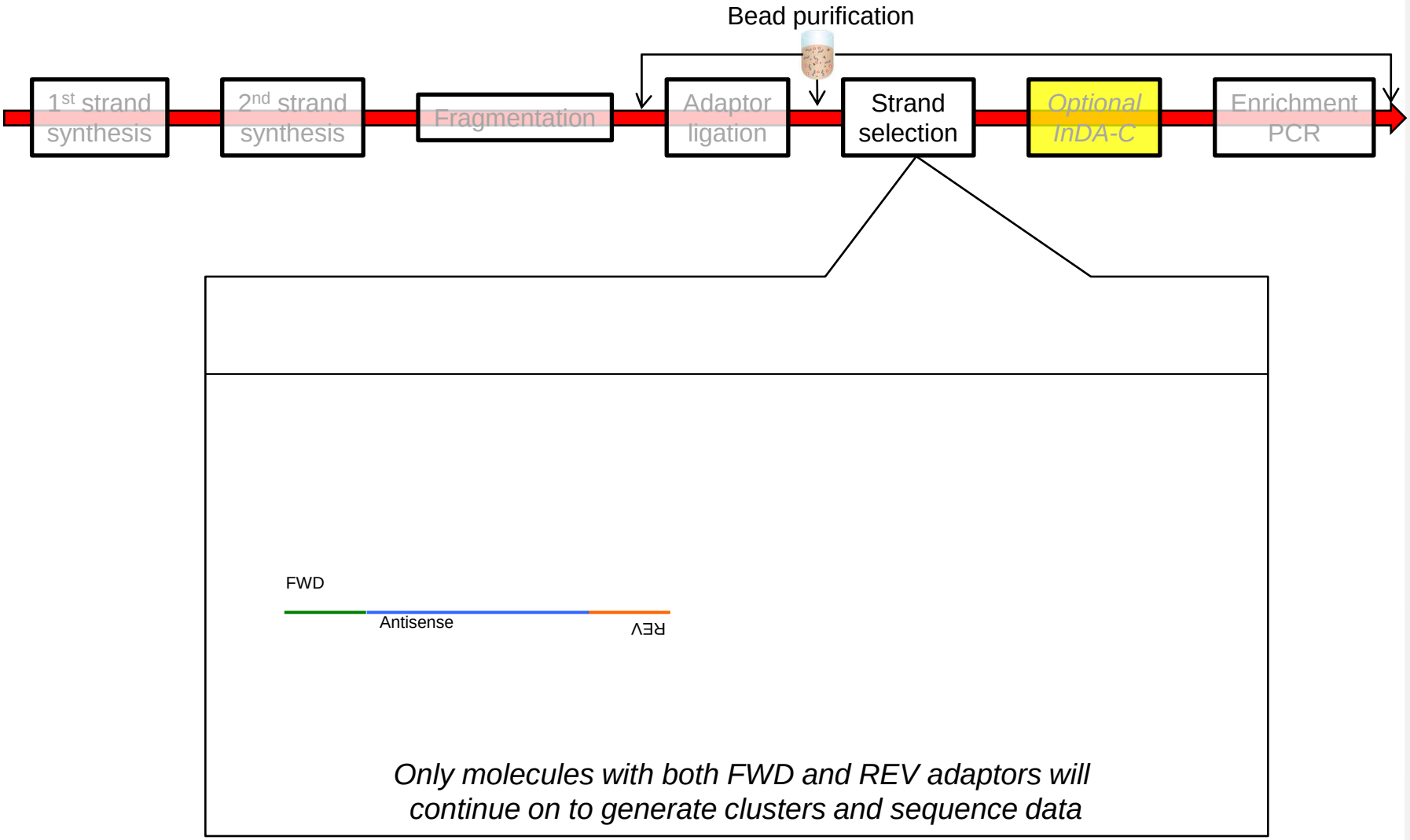
NuGEN imagine more from less®

Strand-Specific RNA-Seq: Strand Selection



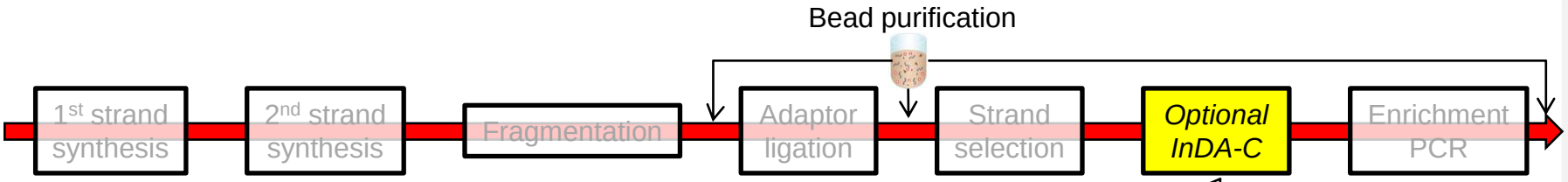
NuGEN imagine more from less®

Strand-specific RNA-Seq: Strand Selection

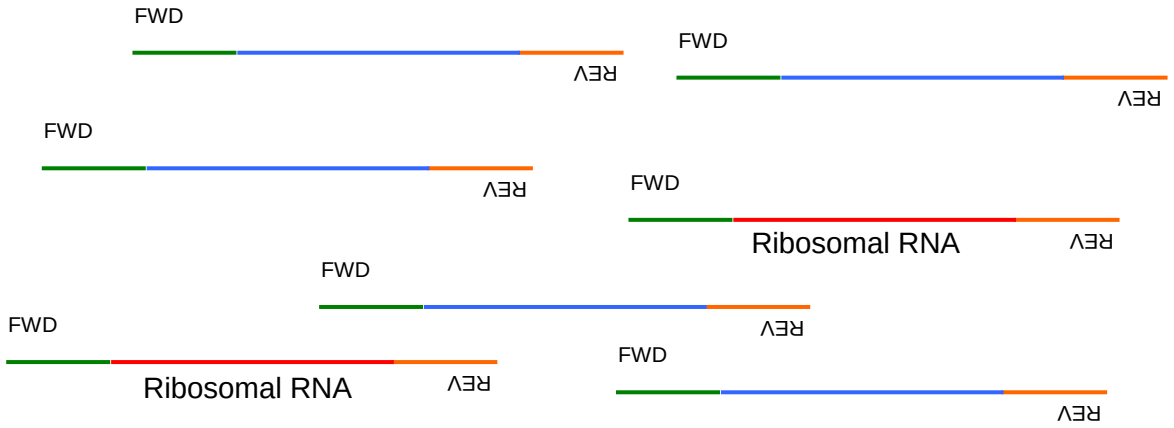


NuGEN imagine more from less®

Strand-Specific RNA-Seq: InDA-C

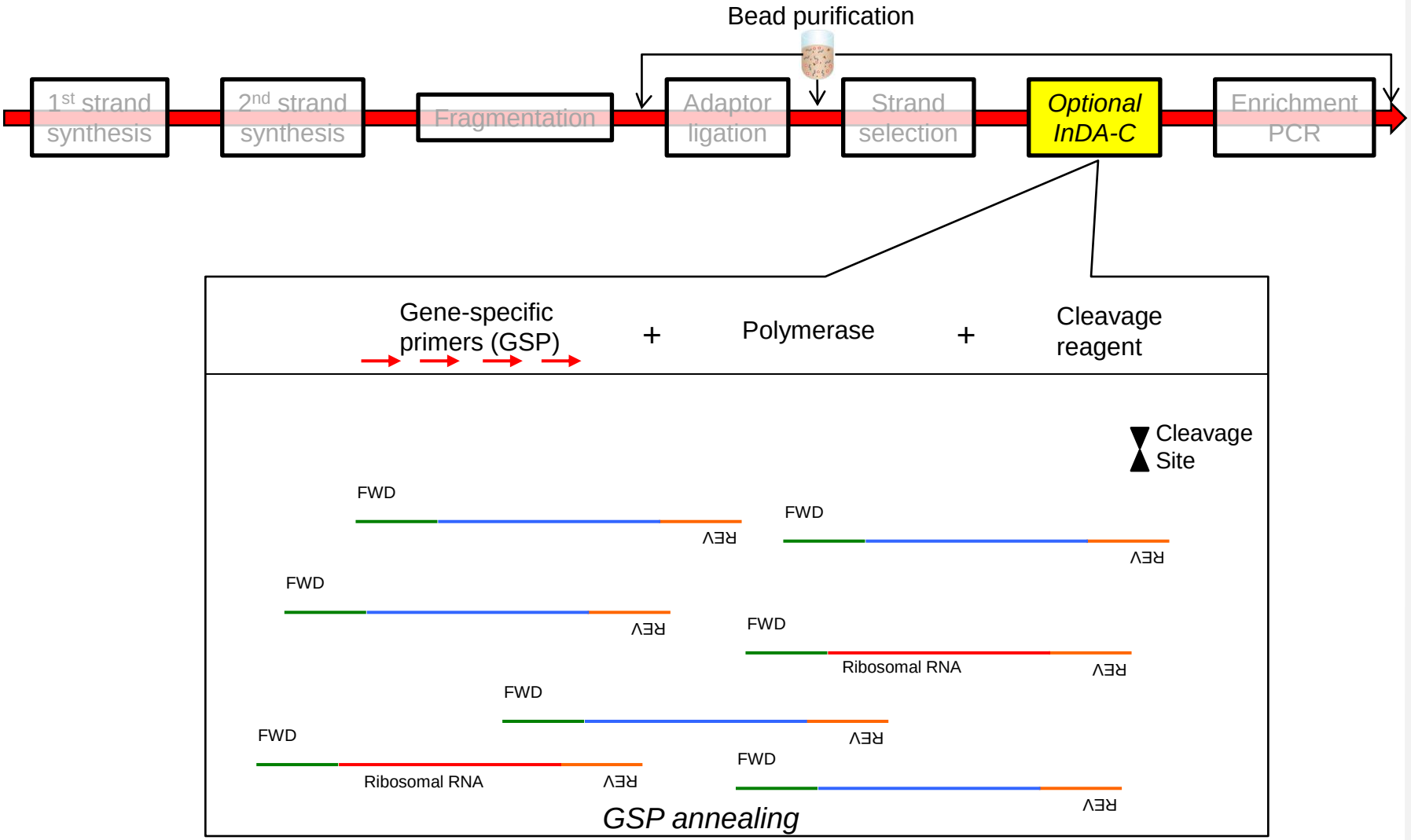


Insert Dependent Adaptor Cleavage (InDA-C)



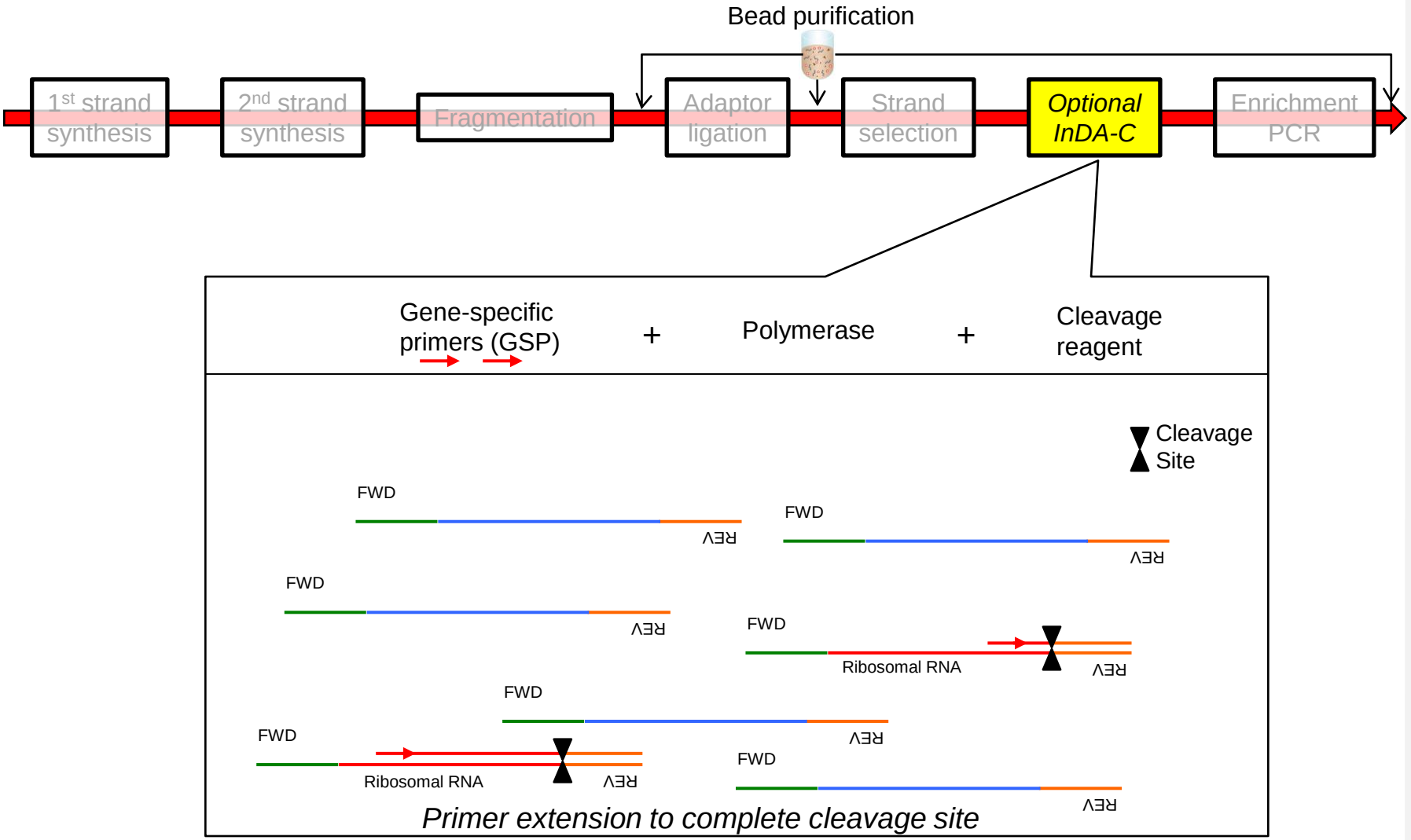
NuGEN imagine more from less®

Strand-Specific RNA-Seq: InDA-C



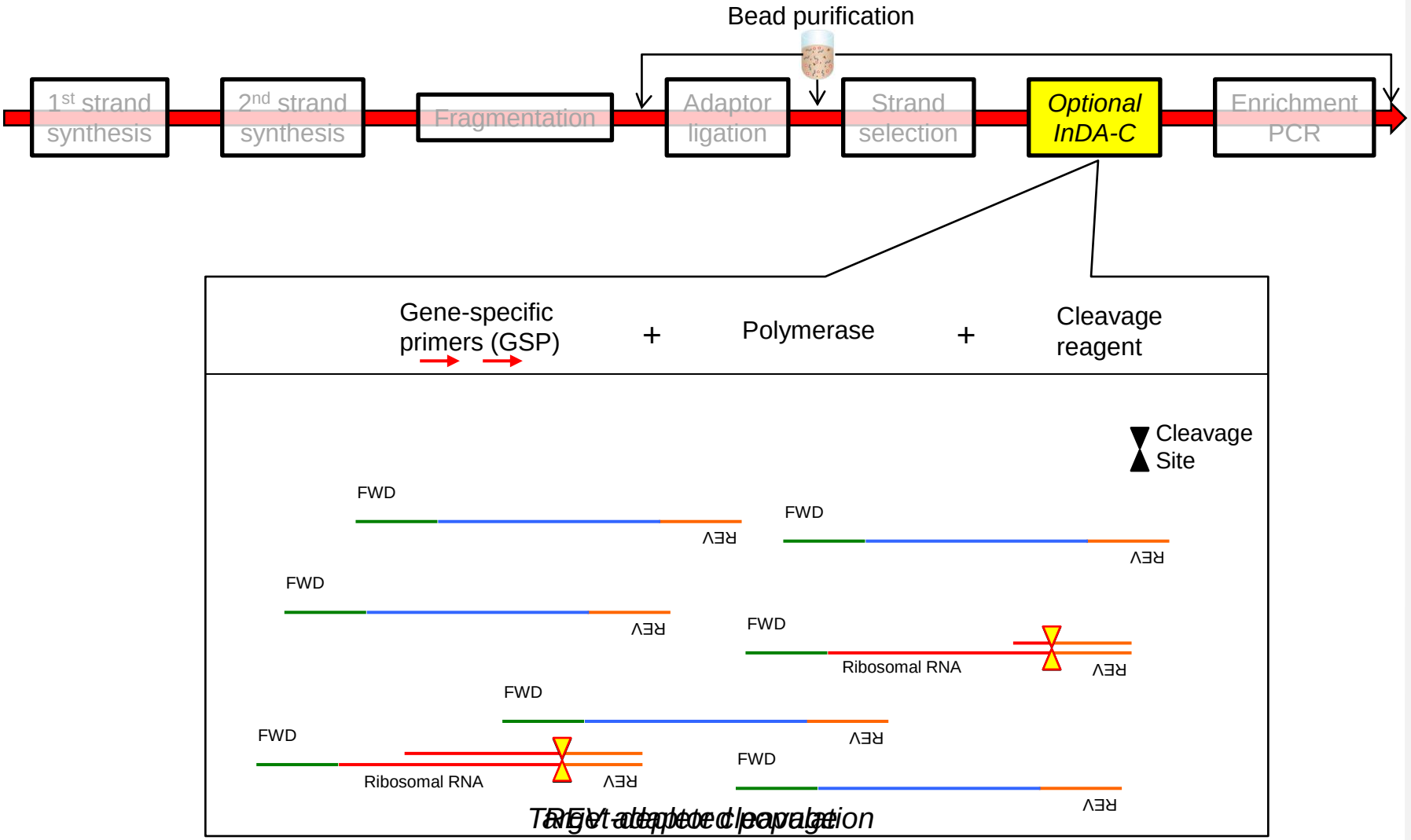
NuGEN imagine more from less®

Strand-Specific RNA-Seq: InDA-C

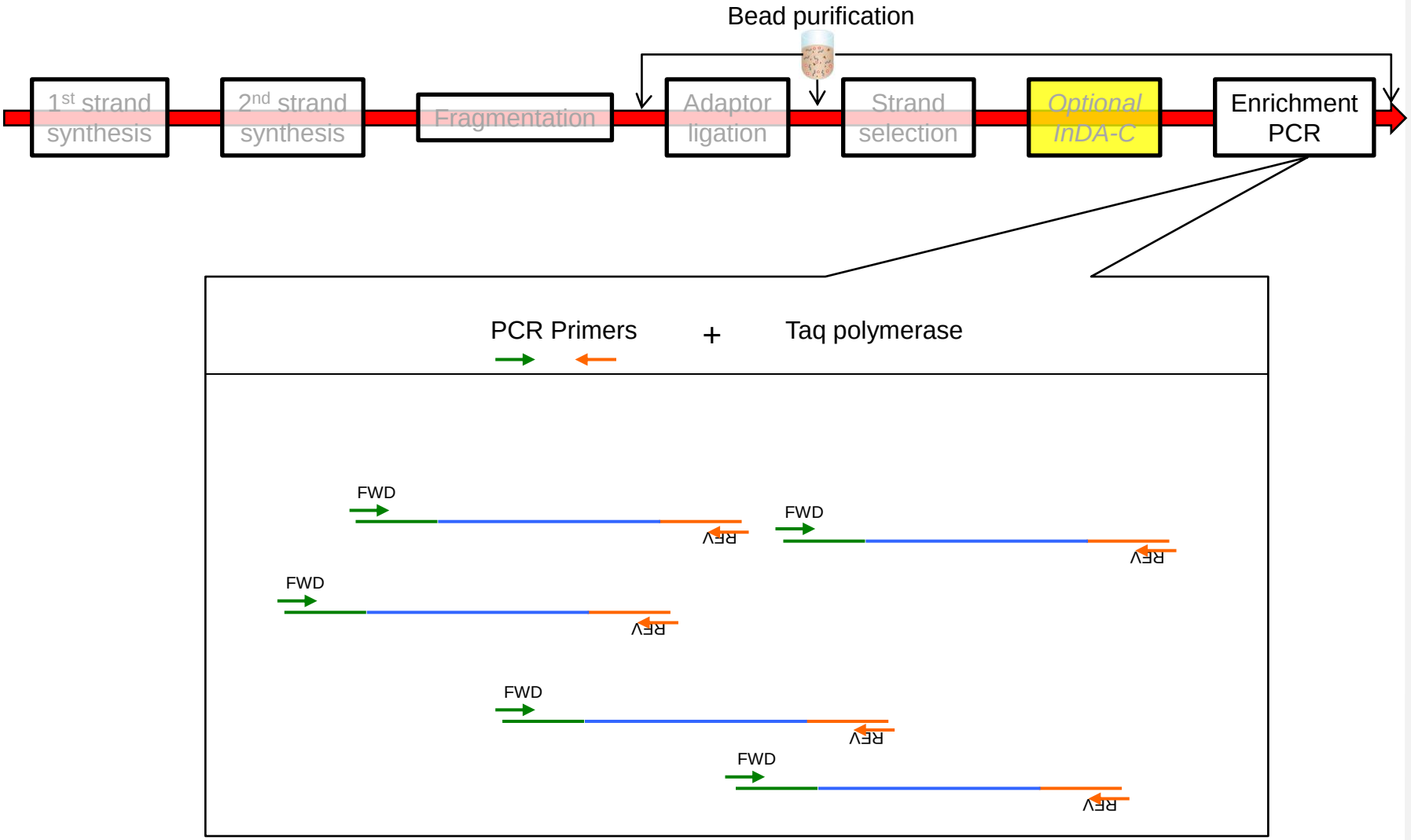


NuGEN imagine more from less®

Strand-Specific RNA-Seq: InDA-C

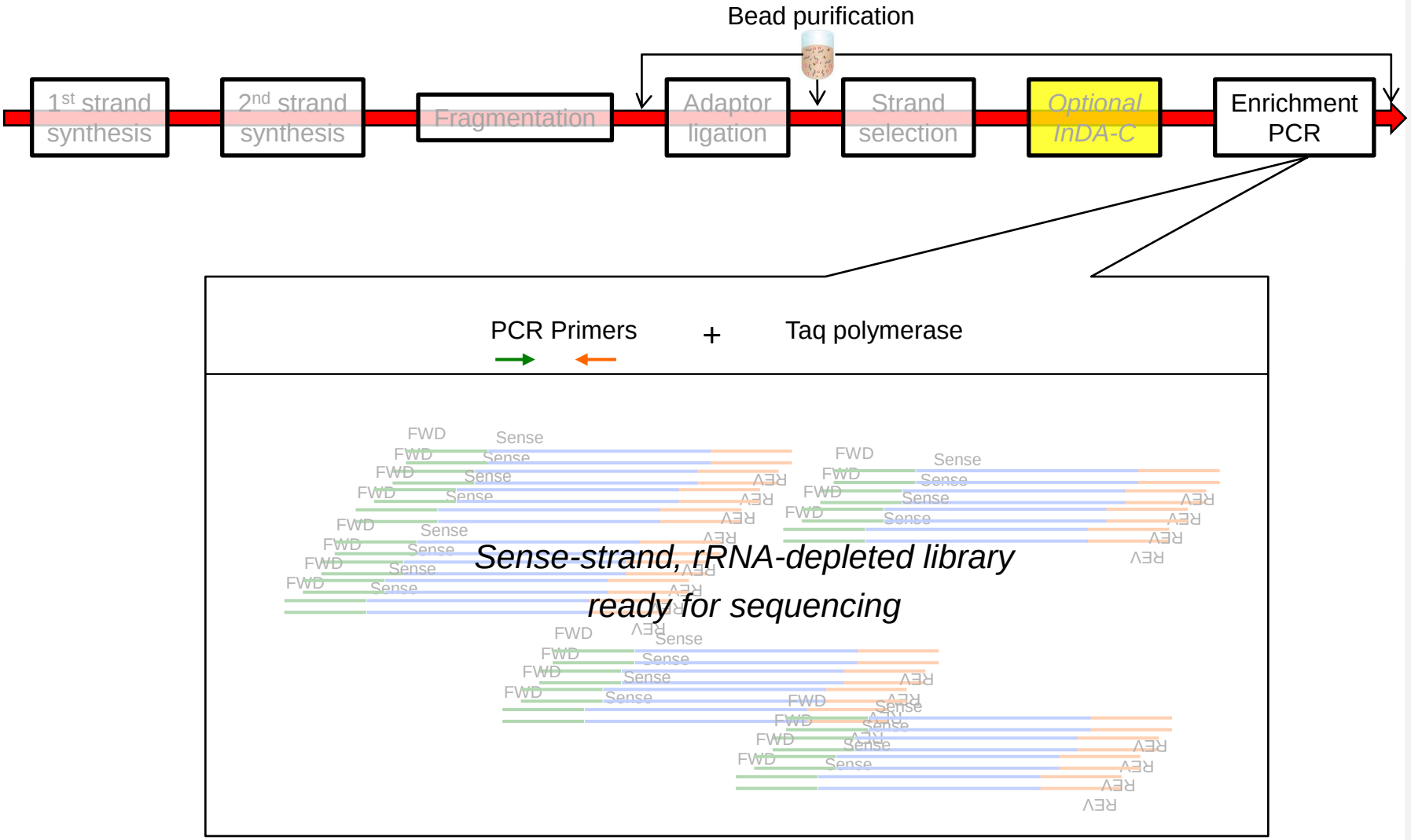


Strand-Specific RNA-Seq: Library Enrichment



NuGEN
imagine more from less®

Strand-Specific RNA-Seq



NuGEN
imagine more from less®

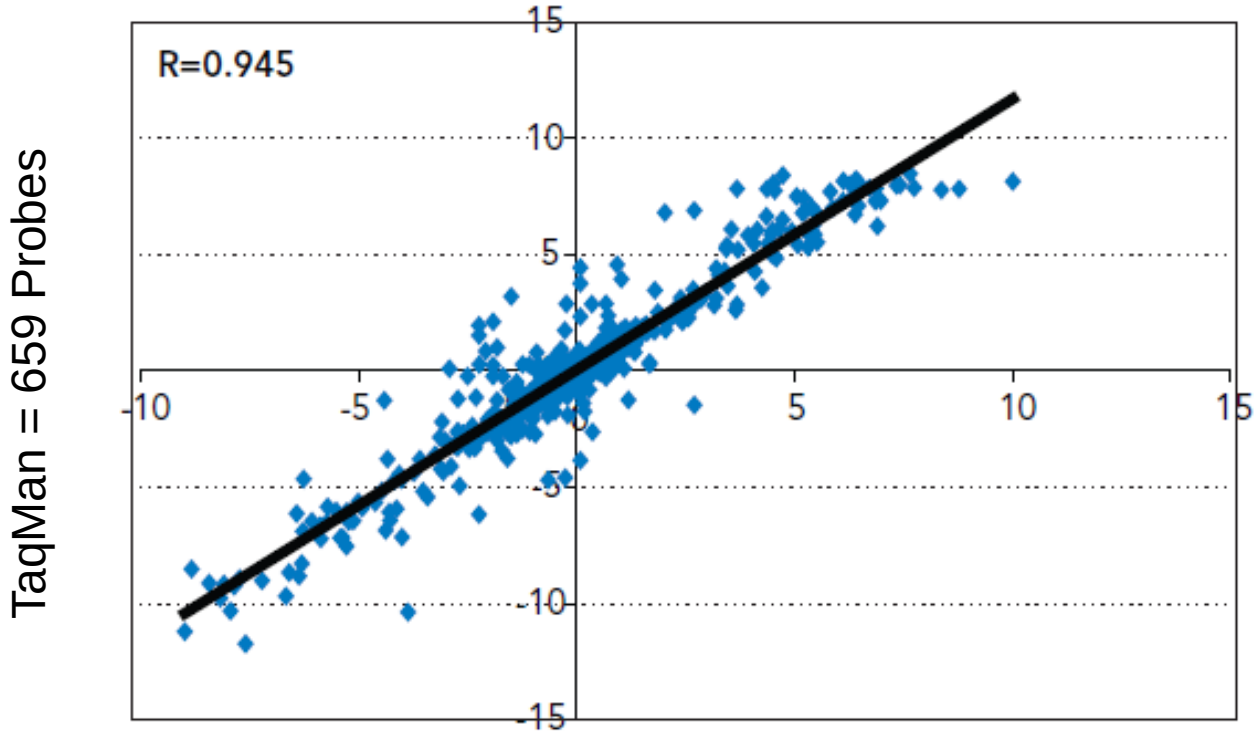
Encore Complete RNA-Seq Sequence Metrics
(Does Not Use InDA-C)

	Source of Total RNA				
	Human (MAQC B)	Human (MAQC A)	Chicken	Mouse	Rat
% of Total Reads					
Not aligned	4%	5%	17%	14%	16%
Aligned	96%	95%	83%	86%	84%
% of Mapped Reads by Category					
All non-rRNA	69.1%	76.8%	62.7%	63.9%	80.1%
Non-rRNA single site	50%	51%	54%	37%	55%
Non-rRNA multiple site	18%	25%	7%	23%	21%
All rRNA	30.9%	23.2%	37.3%	36.1%	19.9%
Mitochondrial rRNA (12S and 16S)	10%	5%	10%	15%	4%
Cytoplasmic rRNA (5.8S, 18S, and 28S)	20%	18%	27%	20%	15%
Distribution of RefSeq Reads					
Exons	44%	38%	33%	49%	42%
Introns	36%	36%	7.0%	20%	16%
Intergenic	16%	24%	60%	32%	42%
RefSeq Strand Retention (% of Reads in Sense Orientation)					
Exons	98.2%	95.7%	98.8%	98.1%	97.8%
5′ UTR	96.3%	96.1%	99.2%	98.6%	98.7%
3′ UTR	96.9%	93.6%	98.2%	96.7%	97.4%

- High degree of alignment
- Reduced ribosomal RNA
- Excellent strand retention

NuGEN
imagine more from less®

Differential Gene Expression vs qPCR

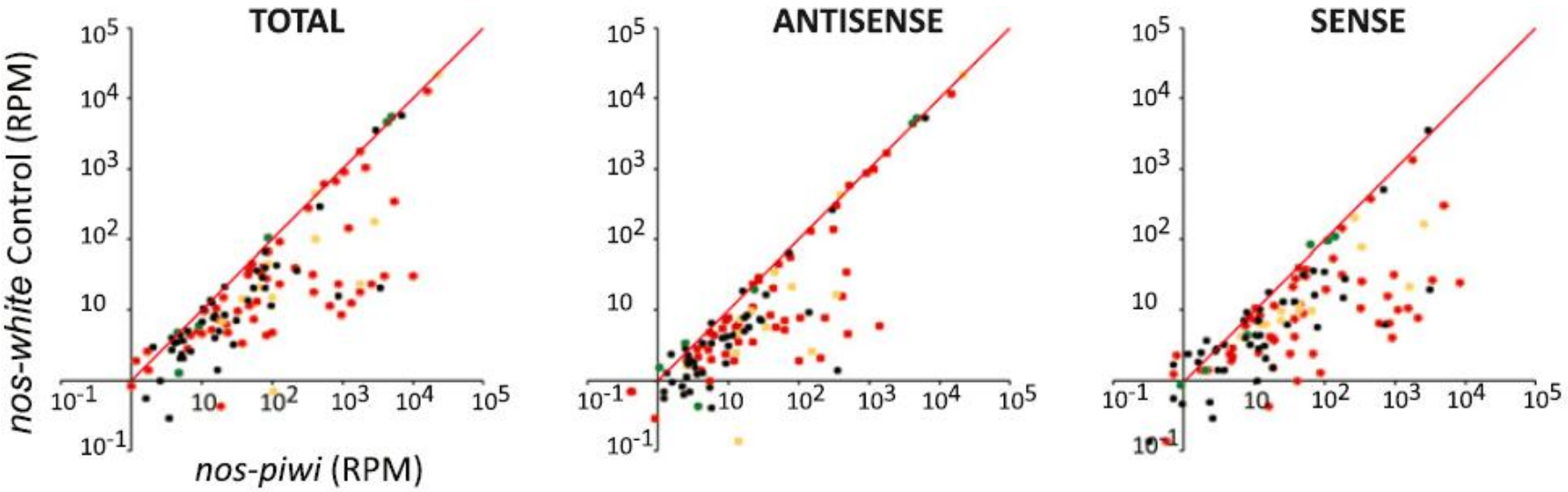


Encore Complete RNA-Seq Library System*

*MAQC log2 RPKM values B (Brain) – log2 A (UHR)

NuGEN imagine more from less®

Impact of Piwi Loss on Steady-State RNA Levels



- Study examines the role of the Piwi family of nucleic acid-binding proteins in silencing of transposons in the *Drosophila* germline cells with strand-specific RNA-Seq
- The results show comparison of RNA levels in control (Y-axis) vs *piwi* knockdown (X-axis) cells using samples prepared with the Encore Complete RNA-Seq Library System
- The data shows altered expression in the *piwi* knockdown cells, with a magnified effect by considering only the sense-oriented reads

From Rozhkov N., Hammell M., and Hannon, G., *Genes & Development*, **27**: Feb. 7, 2013

NuGEN imagine more from less®

Encore Complete Prokaryotic RNA-Seq System

- Uses both Selective Priming and InDA-C probes targeting bacterial sequences to minimize the presence of rRNA in final library
- Universal solution for a broad range of bacterial species and mixed populations
- Strand-specific RNA-Seq using 100 ng total RNA

Sequencing Alignment Metrics for Bacterial cDNA Libraries

	Source of Total RNA					
	S. aureus (33% G+C)		*E. coli* (51% G+C)		*B. thailandensis* (68% G+C)	
	Control	Encore Complete	Control	Encore Complete	Control	Encore Complete
Total Reads	39,482,291	10,226,580	27,161,879	9,355,578	36,471,177	5,196,066
% of Total Reads						
Not Aligned	0.4	3.0	0.7	4.9	0.8	11.4
Aligned	99.6	97.0	99.3	95.1	99.2	88.6
% of Mapped Reads by Category						
Non-rRNA	2.8	34.6	1.0	11.7	2.8	26.6
rRNA	97.2	65.4	99.0	88.3	97.2	73.4
Relative Statistics (Encore Complete vs. Control)						
Non-rRNA Enrich-ment	12.4X		11.7X		9.5X	
rRNA Removed (%)	94.5		92.4		92.1	
Global RPKM (R value)	0.84		0.88		0.83	

NuGEN imagine more from less®

Encore Whole Blood RNA-Seq Library System

- Reticulocyte globin transcripts can constitute 50-70% of total mRNA in whole blood
- NuGEN offers a complete solution for strand-specific RNA-Seq from adult human whole blood total RNA:
  - Compatible with total RNA isolated from PAXgene tubes
  - Significantly reduces the %globin and %rRNA reads in the final library using InDA-C
  - Target input is 100 ng total RNA

NuGEN
imagine more from less®

Human Globin Transcript Abundance

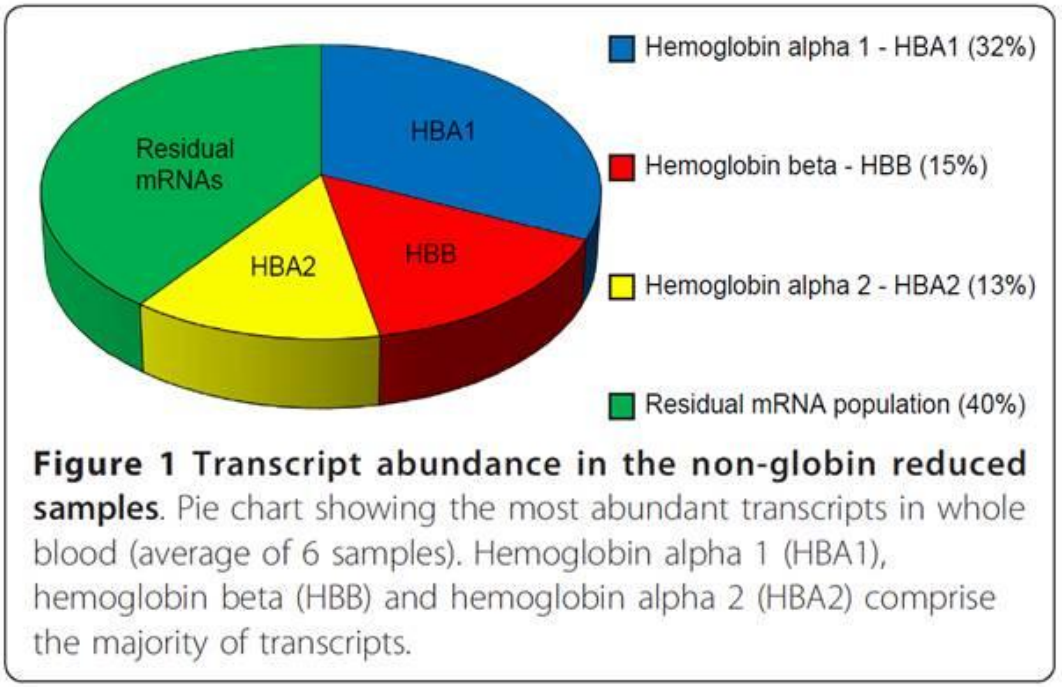


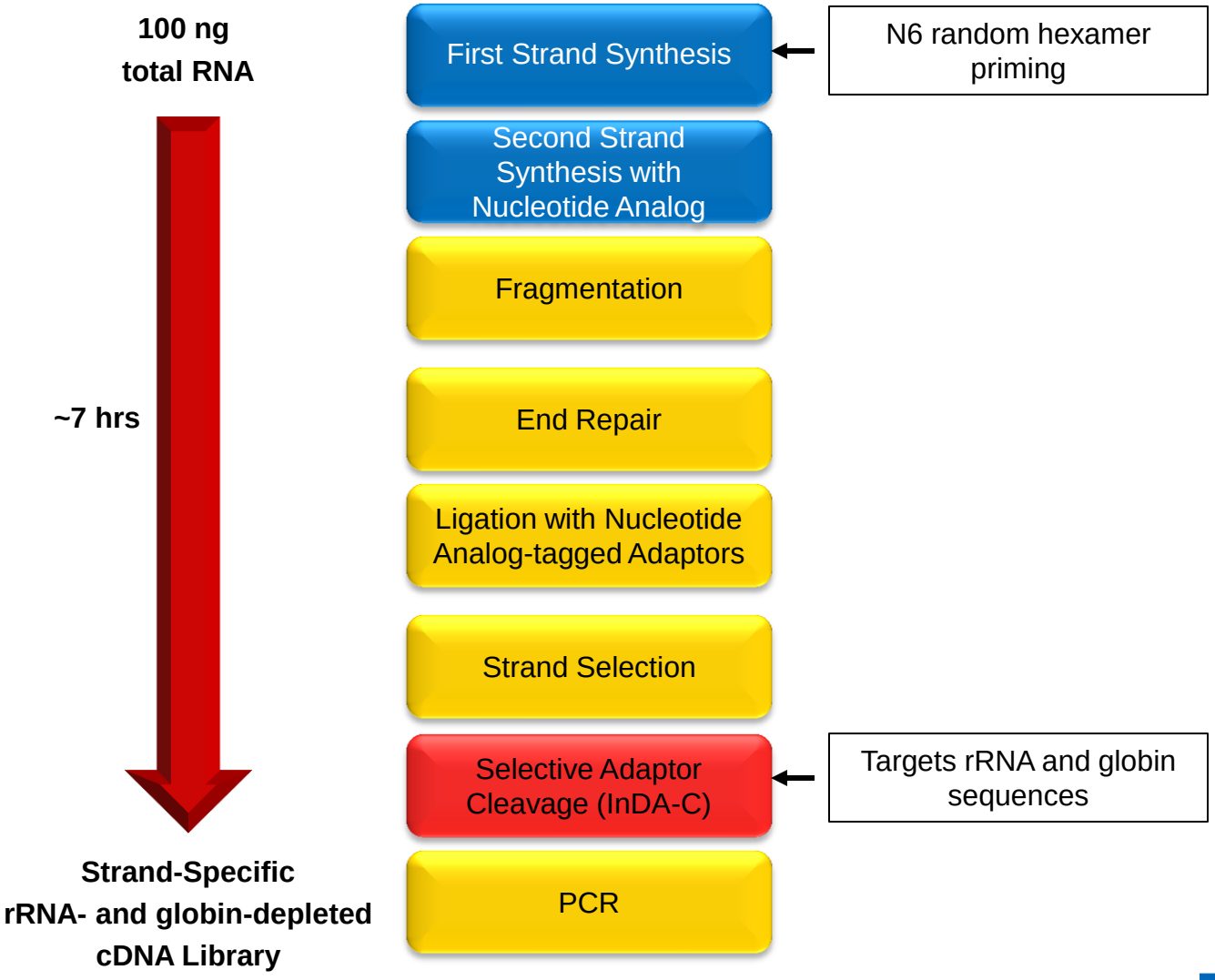
Figure 1 Transcript abundance in the non-globin reduced samples. Pie chart showing the most abundant transcripts in whole blood (average of 6 samples). Hemoglobin alpha 1 (HBA1), hemoglobin beta (HBB) and hemoglobin alpha 2 (HBA2) comprise the majority of transcripts.

InDA-C probes target all four transcript classes

From: Mastrokolias et al. *BMC Genomics* 2012, **13**:28

NuGEN imagine more from less®

Workflow for Whole Blood RNA-Seq



NuGEN
imagine more from less®

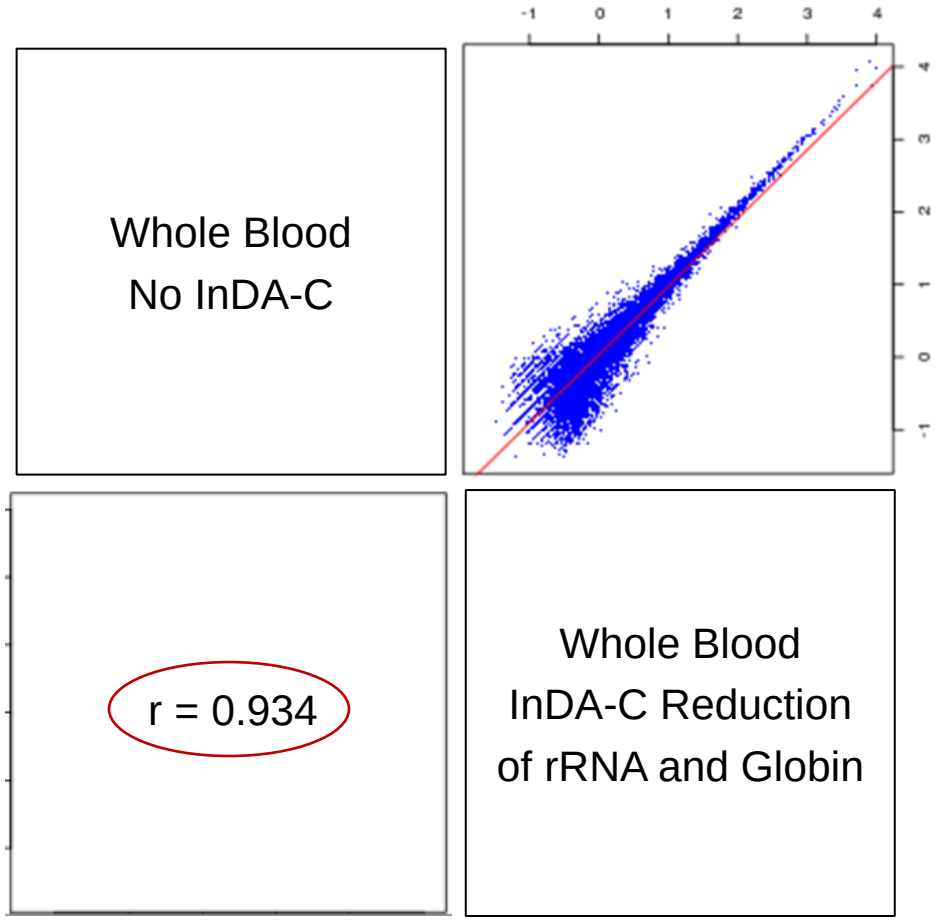
Globin and rRNA Reduction

InDA-C Probes	Genes Targeted	% Aligned	% rRNA Reads	% Globin Reads
rRNA	12S, 16S, 18S, and 28S	92.8	1.6	37.8
Globin	HBA1, HBA2, HBB, and HBD	89.6	9.5	2.7
rRNA + Globin	All of the above	89.3	2.4	3.1

100 ng input of total RNA from adult human whole blood using standard PAXGene collection and isolation protocols; %rRNA and %globin are percentages of PF aligned reads.

InDA-C has Minimal Impact on Non-Targeted Transcripts

- Compare FPKM plots for libraries generated with and without InDA-C treatment
- Reads mapping to rRNA and Globin have been removed from the analysis to show effects on informative transcripts



Ovation Ultralow Library Systems

- **Low input requirements** – Library construction with as little as *1.0 ng DNA*
  - Enables sequencing of low abundance samples without pre-amplification
- **A complete solution for a range of NGS applications** –
  - RNA-Seq
  - Whole genome sequencing
  - Targeted resequencing
  - ChIP-Seq
- **Cost-effective and scalable** – Optional barcoding up to 96-Plex

Human Genome Sequencing

- Goals:
  - Evaluate alignment, coverage, duplications and GC bias at low input of gDNA
  - Demonstrate utility of Mondrian™ SP System to produce complex libraries from low input samples
- Experimental Design:
  - Ovation SP Ultralow Library
    - HapMap sample NA19240 (child in Yuroba trio set) from the 1000 Genomes project
    - 50 ng fragmented gDNA processed using Ovation SP Ultralow Library System on the Mondrian™ SP Workstation, 10 cycles of PCR enrichment
    - Sequencing performed on HiSeq - 100 bp paired-end data
  - Control Library
    - HapMap sample NA18507 (father in Yuroba trio set) from the 1000 Genomes project
    - Standard Illumina paired-end library construction protocol using 1 µg gDNA
- Complete webinar available at www.nugeninc.com

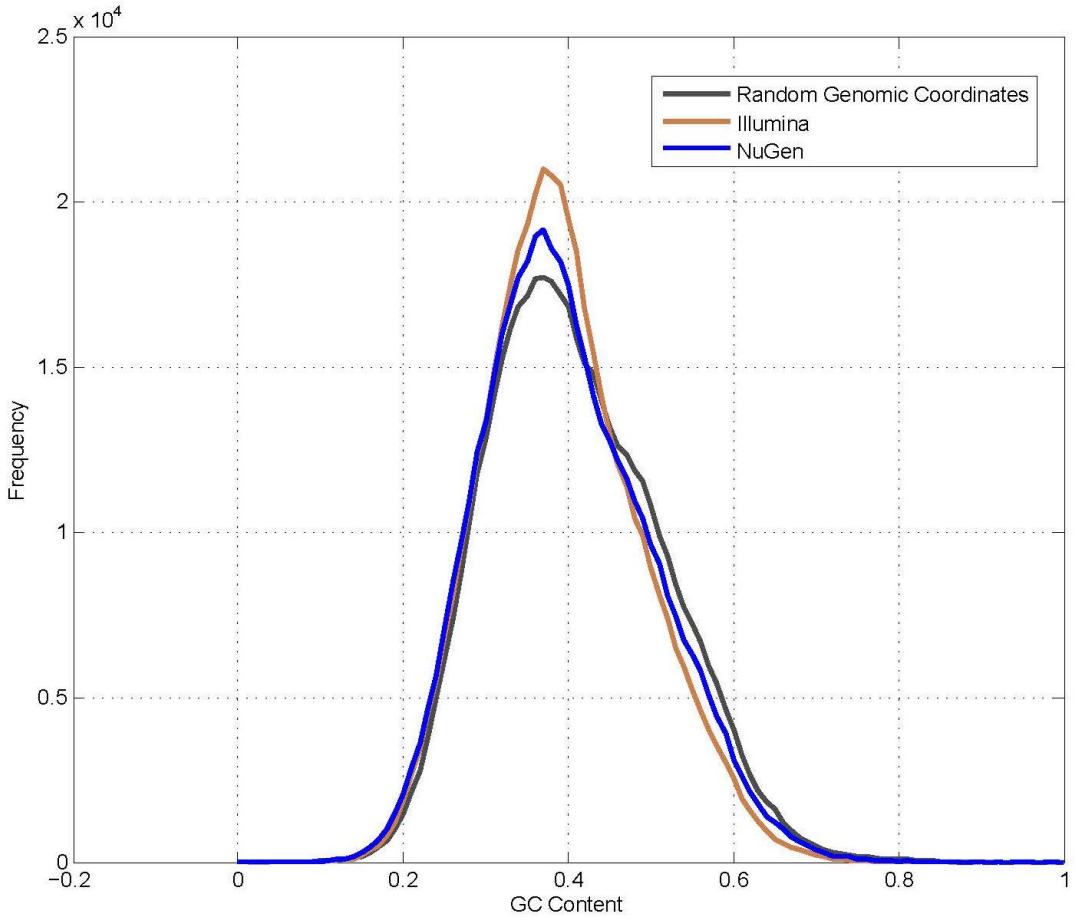
Comparative Alignment Metrics

Metric	Ovation SP Ultralow Library Sample NA19240		Illumina Library Sample NA18507	
	Count	% of Input	Count	% of Input
Total # of input reads	552,195,605	N/A	551,846,550	N/A
Total # reads removed (adapters)	112	0.00%	199,410	0.04%
Total # reads removed (map to more than 1 unique site)	27,668,750	5.01%	31,179,423	5.65%
Total # of reads that align to only 1 site in the genome	487,658,546	88.31%	465,458,145	84.35%
# of reads that did not align to the reference genome sequence	36,868,197	6.68%	55,009,572	9.97%
mapped pairs > 100 kb apart (concatemers)	N/A	0.64%	N/A	0.21%

- Alignment
  - Novoalign using hg19 as reference
  - NA18507 downloaded as fastq files
- NuGEN library shows slightly higher percentage of uniquely aligned reads

NuGEN imagine more from less®

GC Content



Distribution of 1M randomly sampled reads

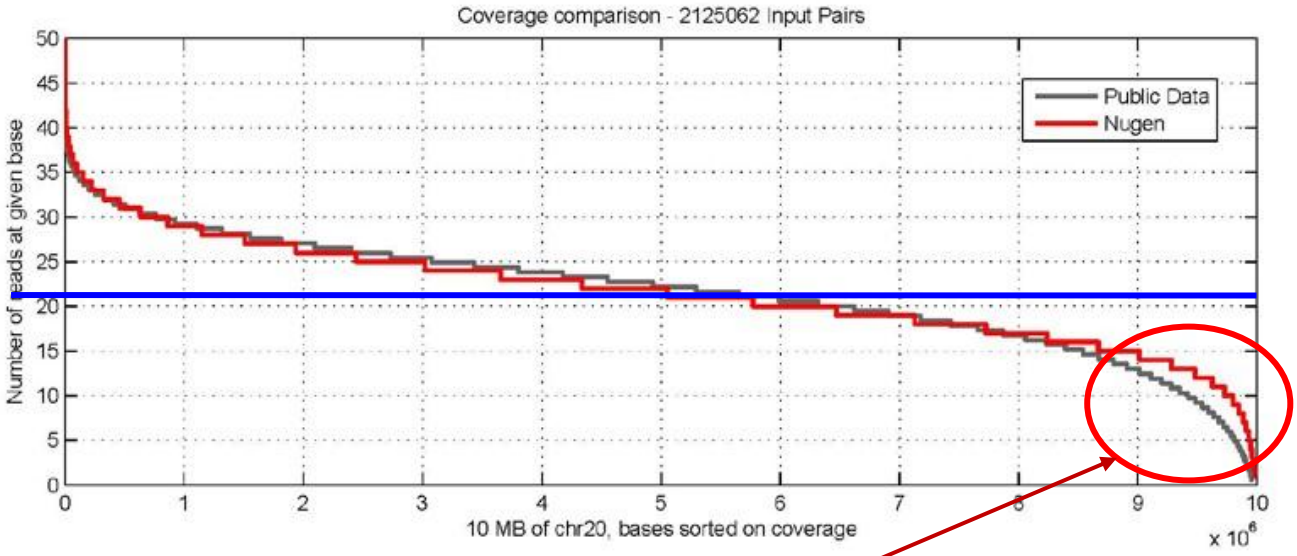
- Human genome is ~ 40% GC.
- GC content distribution of Ovation SP Ultralow library (blue) and Illumina library (orange) plotted relative to predicted GC content profile (gray).
- No significant GC bias observed with either library

NuGEN imagine more from less®

Comparative Coverage Graph

- 10MB of chr20 was chosen to evaluate coverage
- Same number of reads used for both libraries (~21X depth of coverage)

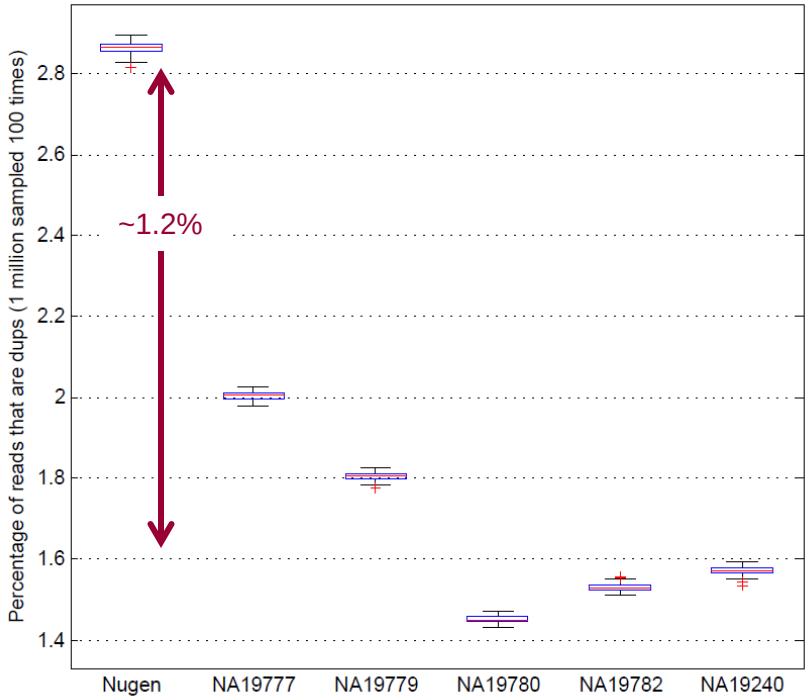
Theoretical perfect coverage distribution



Fewer regions of low coverage compared to standard Illumina library prep using ≥ 1 µg

NuGEN imagine more from less

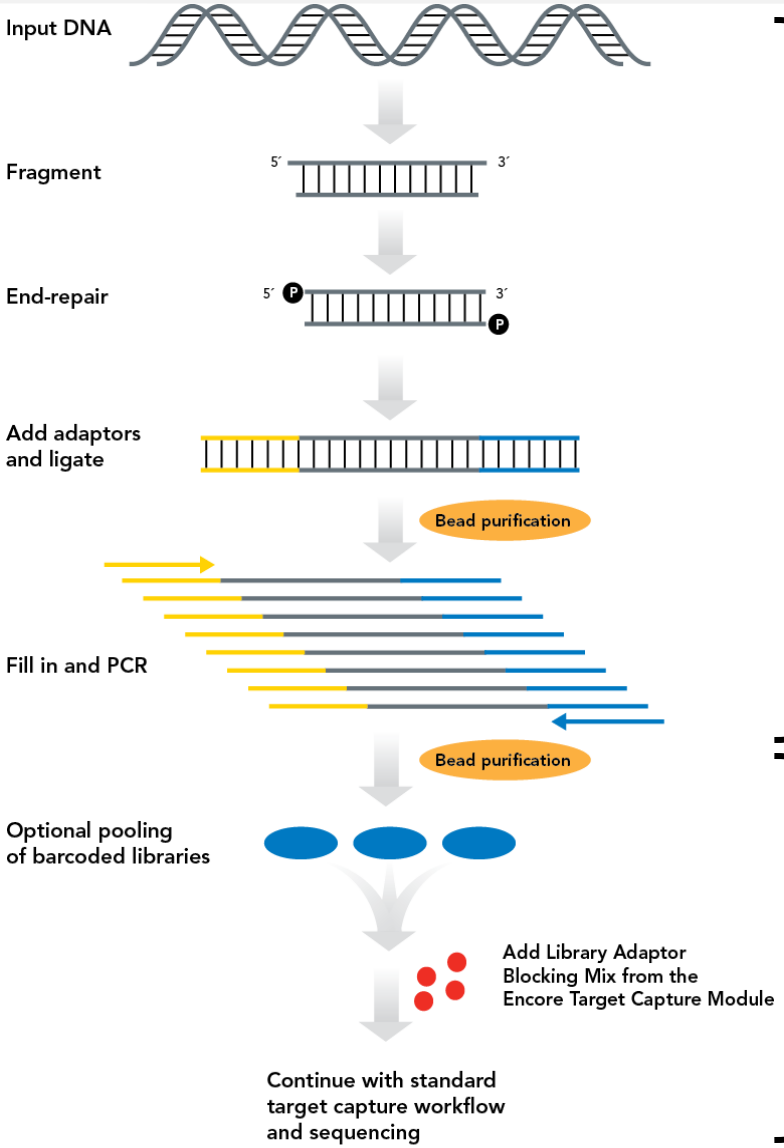
Library Complexity (Duplicate Read Rates)



- Duplicates determined from PE data obtained for chr20 using the same number of input reads (5M)
- Duplicate reads in Ovation SP Ultralow library are elevated but comparable to standard Illumina libraries using much higher input amount - 50 ng vs ≥ 1 µg gDNA input into library prep

NuGEN imagine more from less®

Targeted Resequencing



Ovation Ultralow Workflow

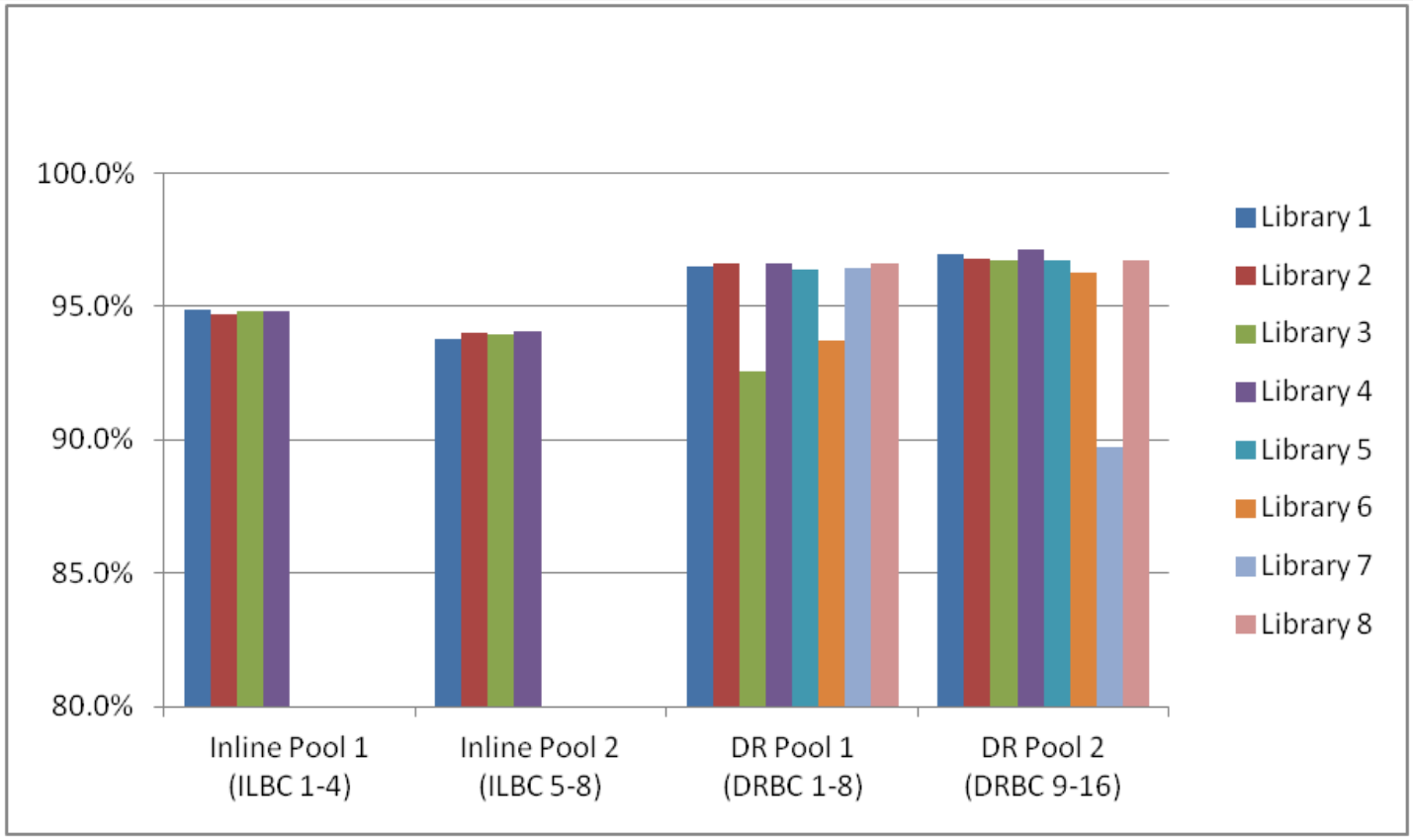
Encore Target Capture Module

NuGEN
imagine more from less®

Benefits for Targeted Resequencing

- **Low DNA input** - Perform target capture with as little as 1.0 ng DNA depending on genome complexity (10 ng for human gDNA)
- **Library pooling before capture** - Pool multiple libraries for input to the same target capture reaction to significantly reduce costs and increase throughput

Multiplexed Target Capture



Target Capture performed with the Agilent Human Kinome SureSelect System with pooled libraries (4 or 8) constructed with 10.0 ng fragmented HapMap hgDNA . The percentage of reads from these libraries that matched the target region (+/- 200 bp) was between 89 and 98 percent.

NuGEN imagine more from less®

Summary

- NuGEN offers a broad range of NGS products, including solutions for
  - Strand-specific RNA-Seq tailored to higher vertebrates, prokaryotes, and human whole blood...more to come
  - Low input library prep using Ovation Ultralow for a range of NGS applications
  - Automation of several NGS workflows using Mondrian SP Workstation

THANK YOU!