JOINT GENOME INSTITUTE
UNITED STATES DEPARTMENT OF ENERGY

Sequencing Technologies: Overview & Capabilities

Chris Daum
Genomic Technologies Workshop
March 22, 2016
cgdaum@LBL.gov

Sequencing Technologies Group

Project Management Office (PMO)

- Define Project
- Obtain Metadata
- Product catalog

work order & flow

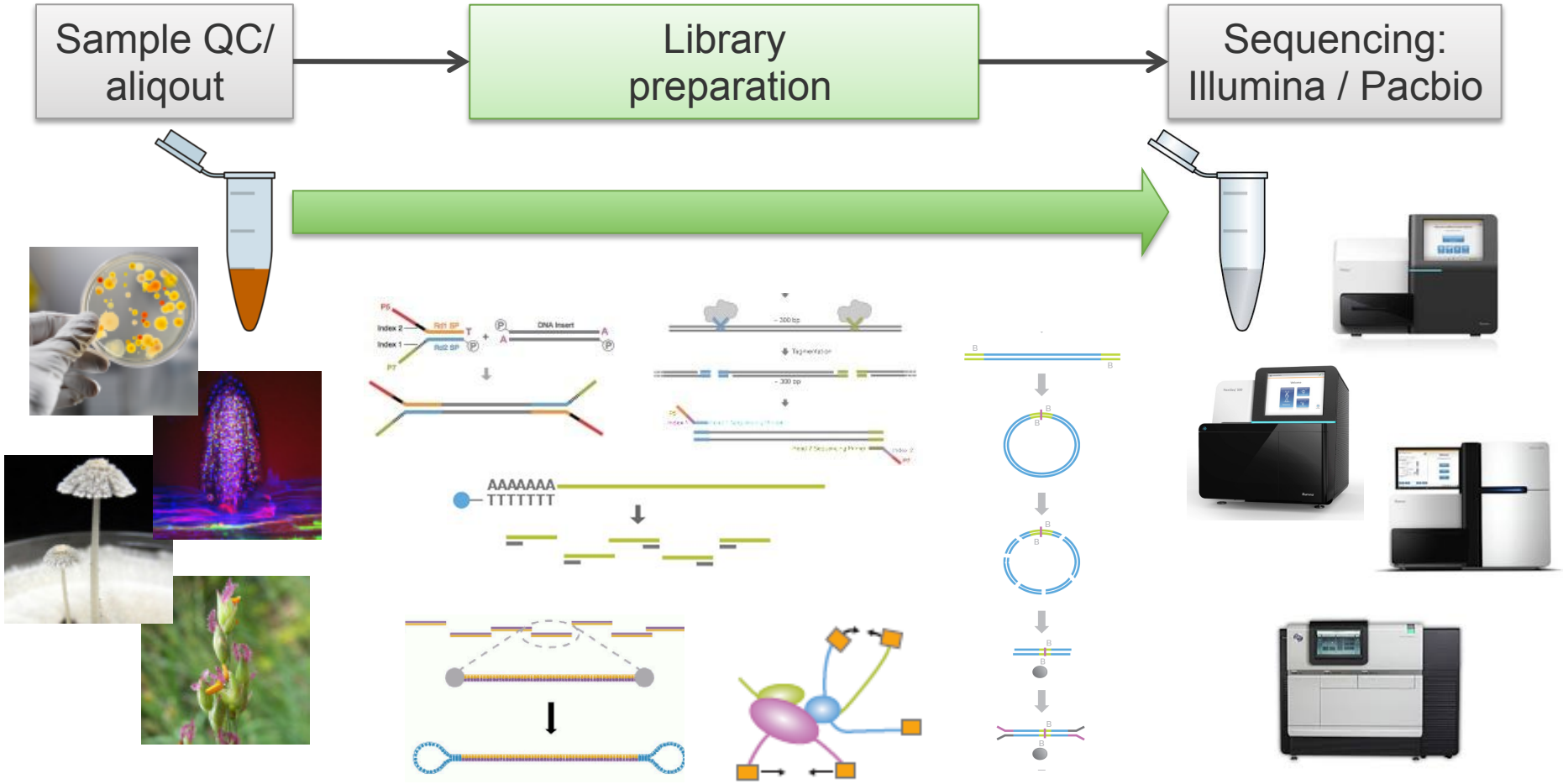
Sequencing Technology ($15M)

Sample Management	Library Construction	Library QC, qPCR, pooling	Sequencing
(5 FTEs)	(11 FTEs)	(3 FTEs)	(5 FTEs)

Area Leads (3); Resource & NTI manager (1); R&D Project manager (1)

- **Streamlined Process from Sample In to Sequence Data Out**
- **Perform Process Optimization & Development:**
  - **New Preps, Applications, Sequencing Technologies**
  - **Continuous Improvement & Lean Manufacturing Six-Sigma**

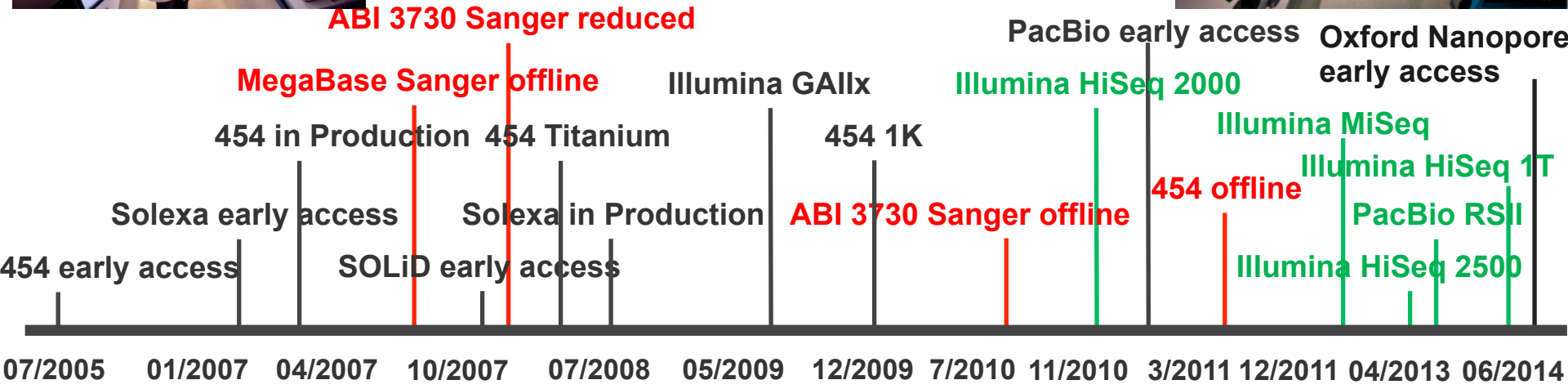
Sequencing Technologies: Sample to Data

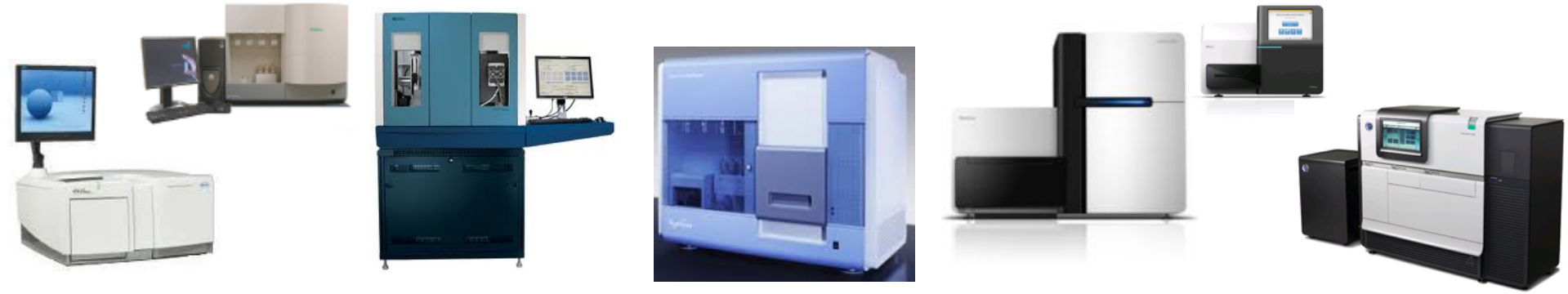


Supported by ITS-GLS LIMS system

Staying State of the Art

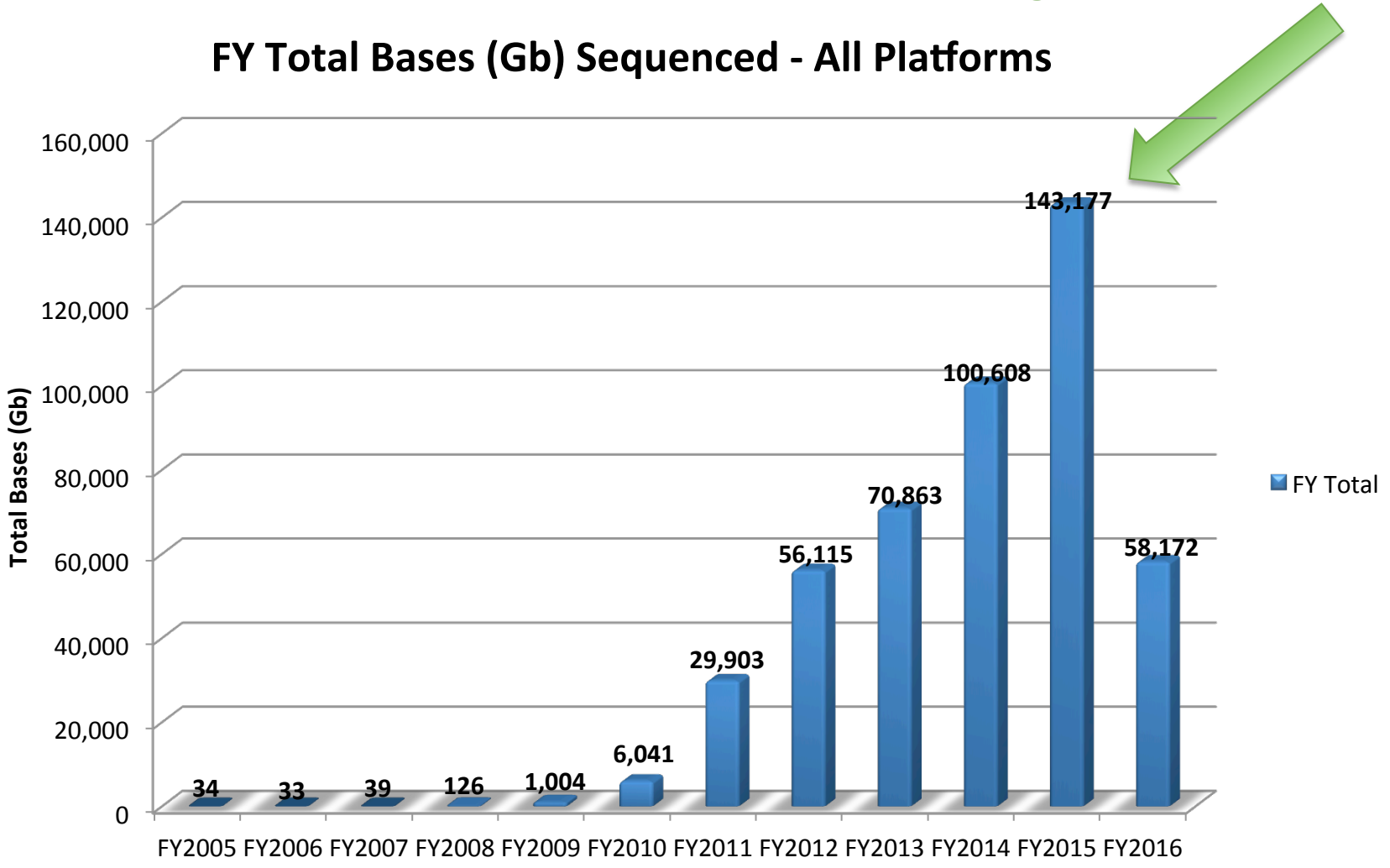
Sanger Sequencing to Next-Gen Sequencing by Synthesis





JGI Yearly Sequence Output

44,000 human genome equivalents!



As of March 21, 2016

JGI Sequencing Platforms Portfolio





	Illumina HiSeq 1T	Illumina HiSeq 2500	Illumina HiSeq 2000	Illumina NextSeq 500	Illumina MiSeq	PacBio RSII	PacBio Sequel
Units	**3**	**3**	**2**	**1**	**5**	**3**	**2**
Reads (Single-Read	>1,500 Million per Flowcell	200 Million per Flowcell	>1,000 Million per Flowcell	400 Million per Flowcell	>10 Million per Flowcell	0.06 Million per SMRT Cell	0.4 Million per SMRT Cell
Readlength	2 X 150bp Max*	2 X 250bp Max	2 X 150bp Max*	2 X 150 Max	2 X 300bp Max	>12,500bp Avg; >40,000bp Max	12,000bp Avg; >40,000bp Max
Total Bases	500 Gb per Flowcell	130 Gb per Flowcell	350 Gb per Flowcell	>100 Gb per Flowcell	5-20 Gb per Flowcell	>0.4 Gb per SMRT Cell for 2hr runs; >0.8Gb for 4hr runs	>2.7 Gb per SMRT Cell for 2hr runs; >5.0 Gb for 4hr runs
Run Time	7 Days for 2 X 150	4.5 Days for 2 X 250	16 Days for 2 X 150	1 Days for 2 X 150	2 Days for 2 X 300	0.08-0.12 Days (2-4 hours)	0.08-0.12 Days (2-4 hours)
Applications	Primary Sequence Generator at JGI	Rapid output HiSeq	Supplement / Backup Platform	Rapid mid-range output; Single Cell	16S/18S iTags, Library QC, R&D	Assembly improvement, de novo, SynBio validation, methylation/ epigenetics	Early Access Testing

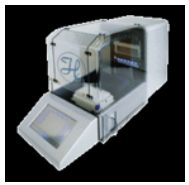
Major Investment in Automation: Process

JGI JOINT GENOME INSTITUTE

Sample Archive & Retrieval → **Sample QC & Aliquot** → **Library Construction** → **Library QC, qPCR, pooling**

SAM

Starlet x 2



Automated Decapper

BioMicroLab Volume Check

BioTek Synergy H1 Plate Reader

Sciclone G3 x3

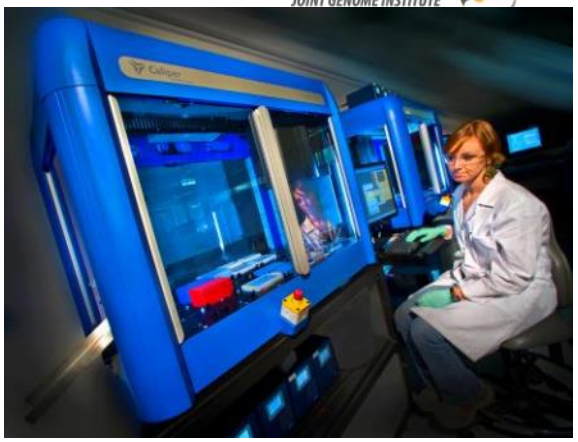
Star x 2

5ul PCR

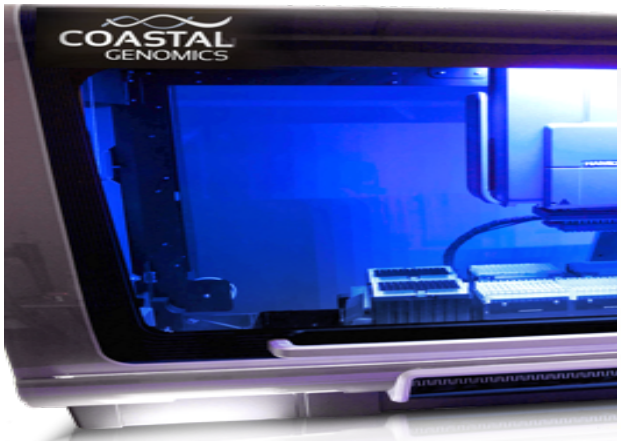
Labcyte Echo

Maximize Consistency, Throughput and Reliability

Automated Library Creation



- **Plate based Automated library preps:**
  - Implemented into Production in late 2011
  - 3 PerkinElmer Sciclone NGS robots
  - 12 production sample prep methods are supported
    - >24,000 sample sequencing libraries prepared in last year
- **Supporting Equipment:**



Sciclone Automated Preps

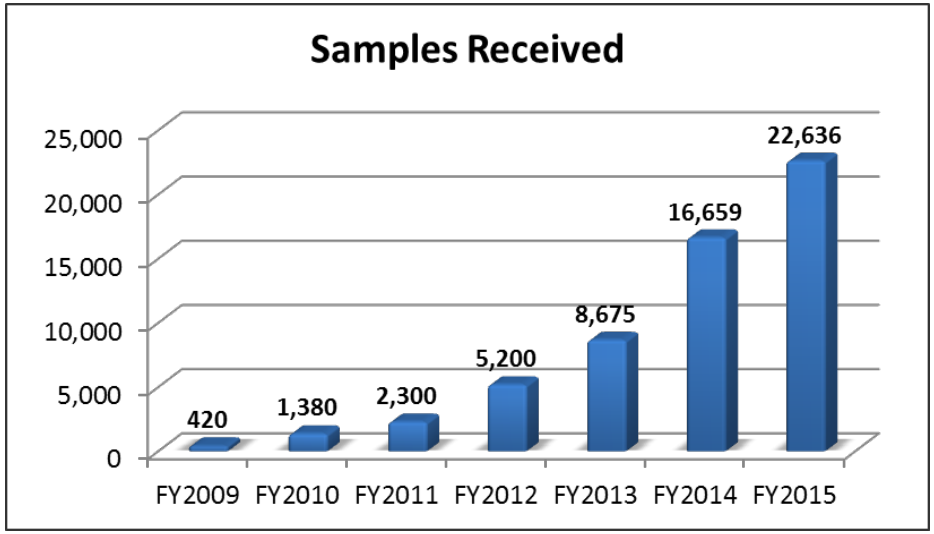
Supported Preps

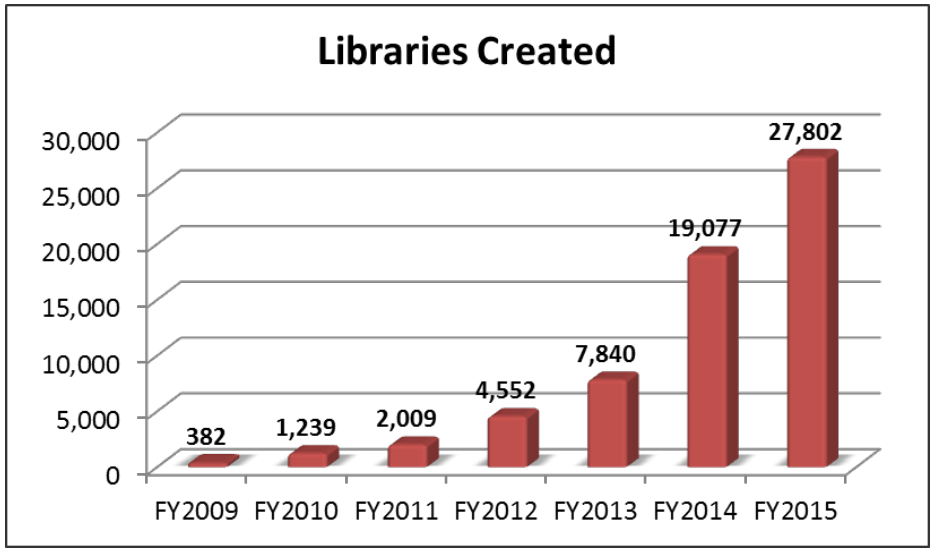
- Illumina gDNA PCR-free WGS Fragment Library prep
- Illumina TruSeq RNA-seq stranded library preps:
  - PolyA selection of mRNA for eukaryotes
  - rRNA depletion for microbes & metatranscriptomes
- Illumina small RNA & miRNA for eukaryotes
- Illumina iTags (16S proks; 18S euks; Fungal ITS)
- Illumina Exome Capture (NimbleGen SeqCap) prep for targeted resequencing
- Illumina NexteraXT prep
- PacBio DNA 2kb libs, and >10kb libs with enzymatic shearing

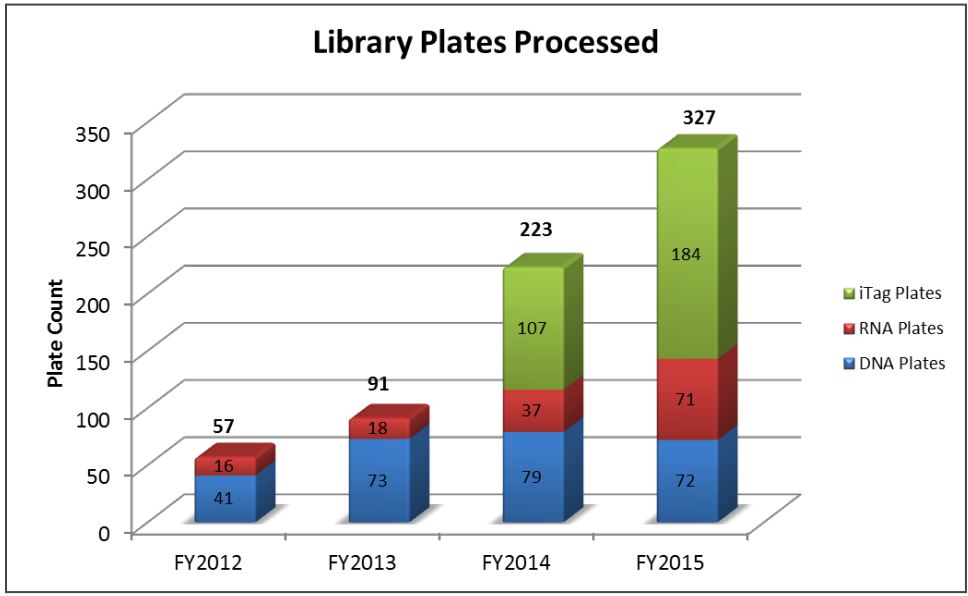
Preps in Development

- Illumina Methyl-Seq (bisulfite conversion) prep
- Illumina 3' RNAseq prep

Continued Growth in Number of Samples Handled





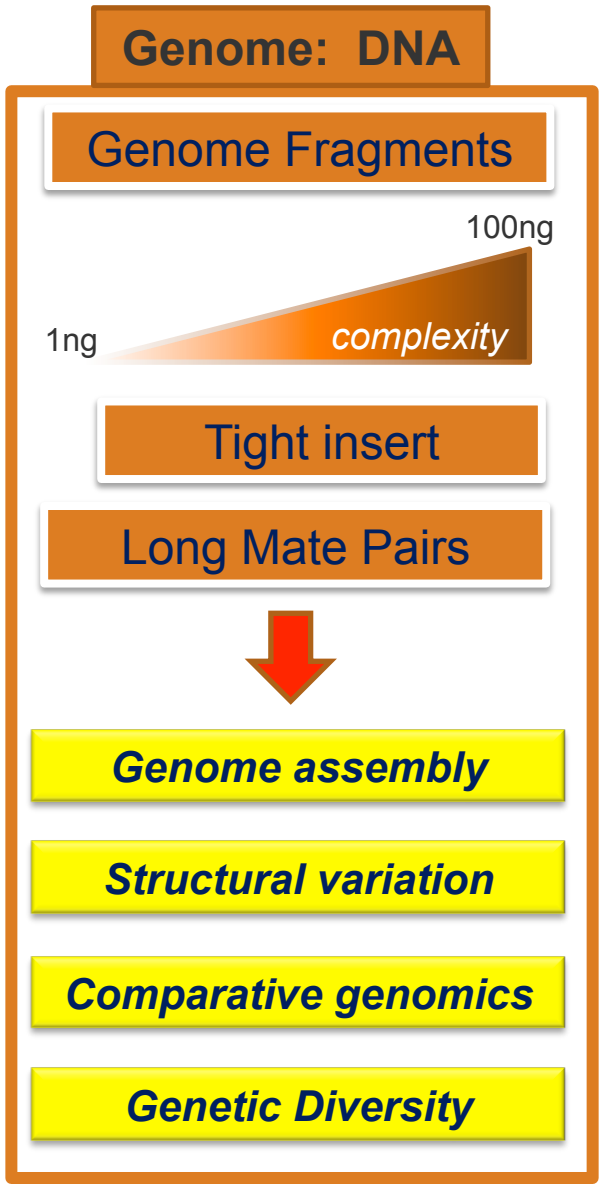


A Growing Portfolio of Library Capabilities

JGI JOINT GENOME INSTITUTE

Genome: DNA	Transcriptome: RNA	Epigenome: DNA-protein
WGS	Stranded RNA-seq	Bisulfite-seq
WGS: Tight insert	Poly A	FAIRE-seq
LMP	rRNA depletion	ChIP-seq
Pacbio long reads	FFPE/LCM	ChIA-PET
↓	smRNA	↓
Genome assembly	Iso-seq	Methylation
Structural variation	↓	Gene regulation
Comparative genomics	Gene annotation	Chromatin conformation
Genotyping	Gene expression	Protein binding

A Growing Portfolio of Library Capabilities



Example: DNA Products

- 270bp
- 500bp
- Low input- Nextera

Genome Fragments

- Tight insert 400bp
- Tight insert 800bp

Genome Fragments

- Pacbio 2kb
- Pacbio 3kb
- Pacbio >10kb
  - Low input 10kb
- Pacbio 20kb
- Oxford Nanopore

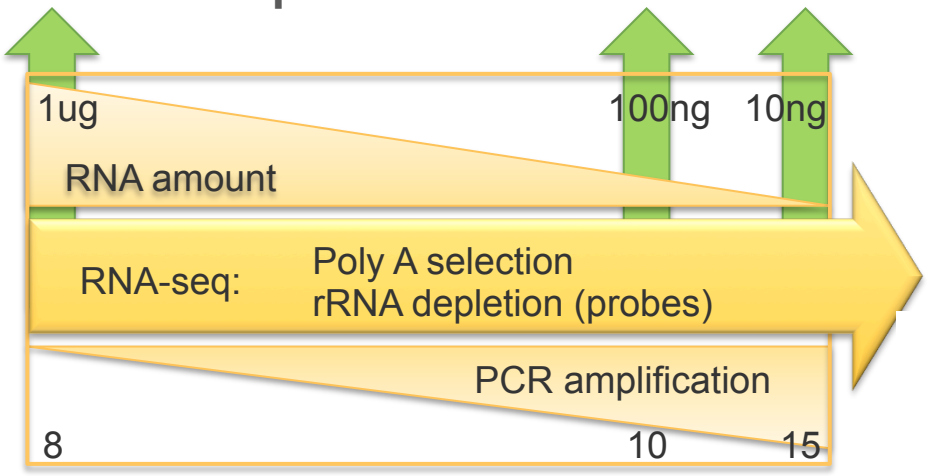
Long Reads

- 2.5kb
- 4kb
- 8kb

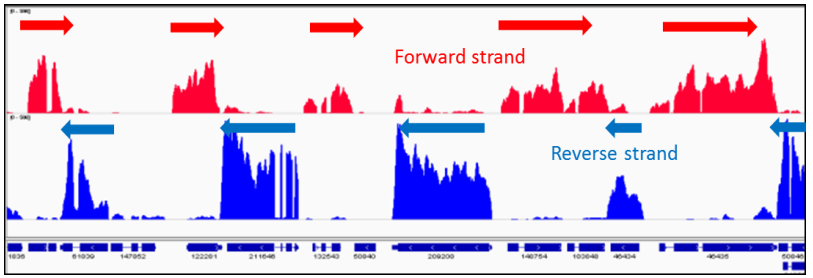
Long Mate Pairs

RNA Sample Prep: Gene expression

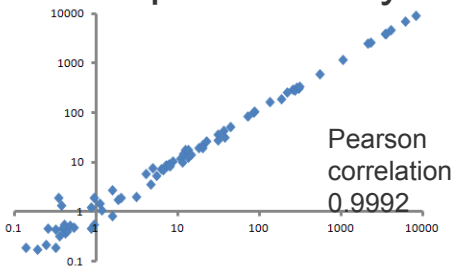
- **RNA-seq**



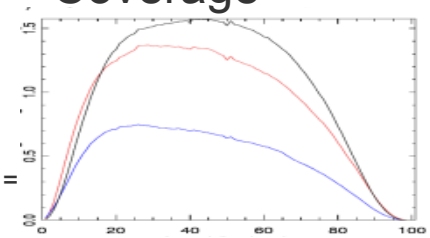
Strandedness



Reproducibility

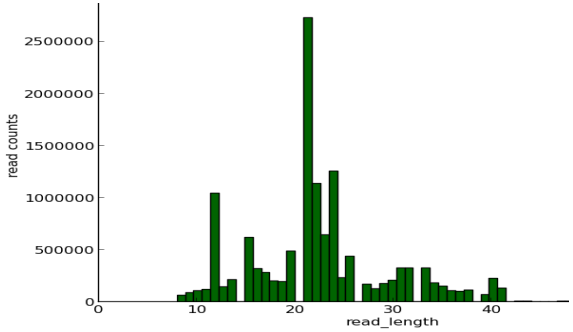


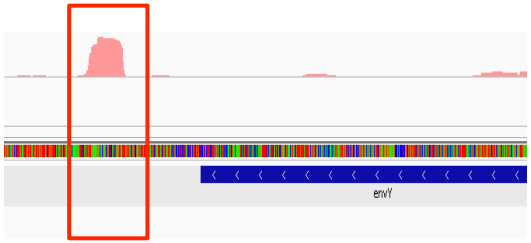
Coverage



- **Small RNA-seq**

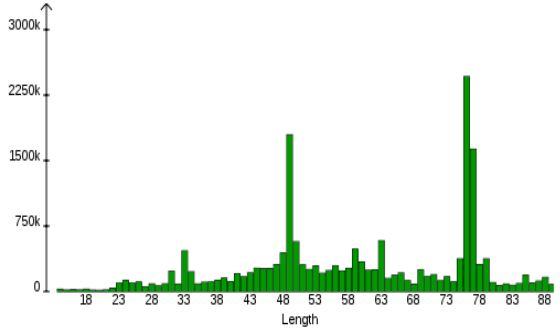
Eukaryotes: 20-40bp





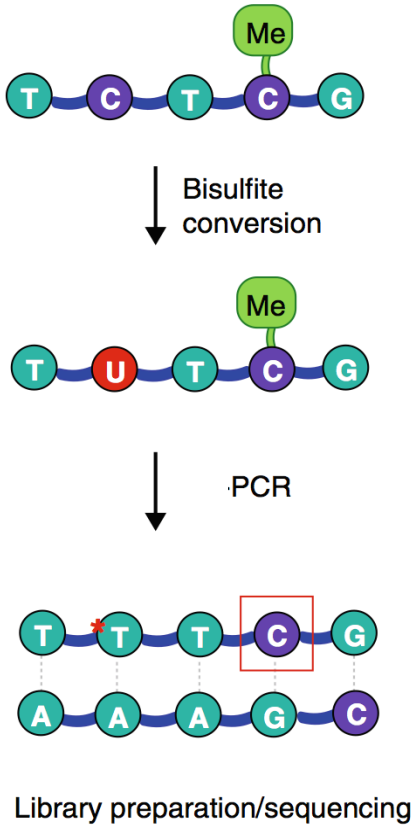
Novel miRNA

Prokaryotes: 50-150bp

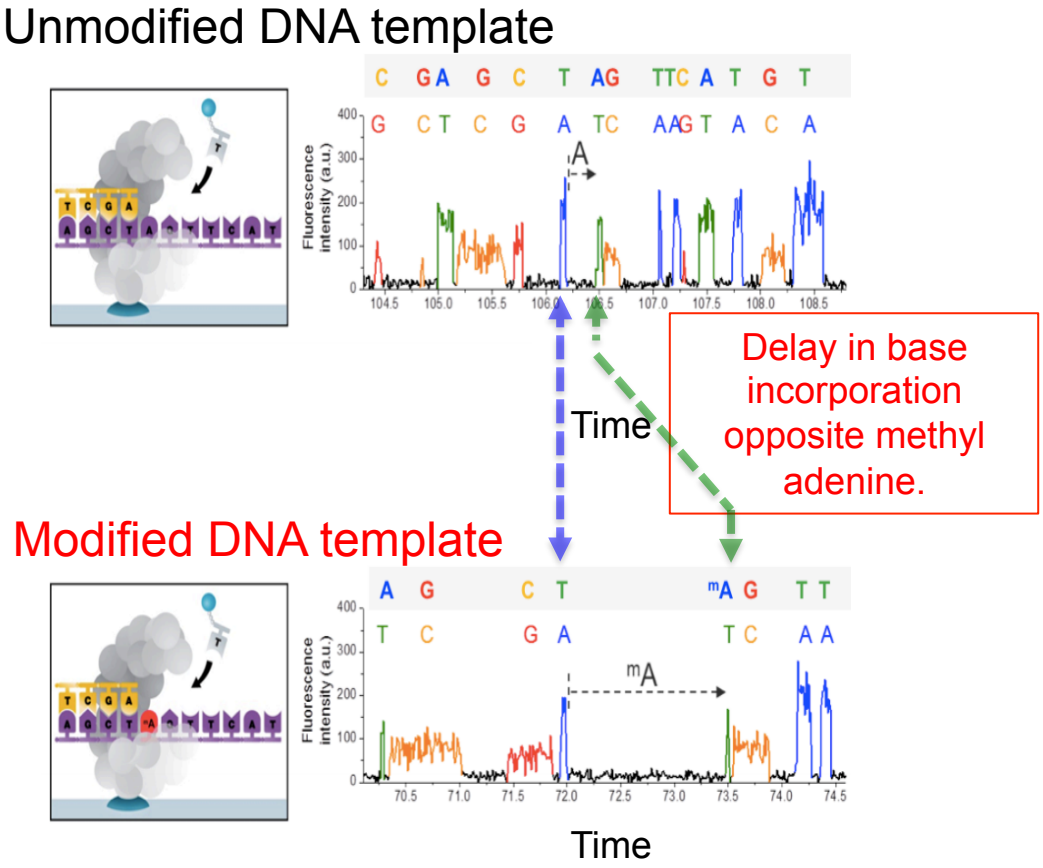


Epigenomic: DNA modifications

5 Methyl C
Bisulfite sequencing



Methyl A
Pacbio sequencing

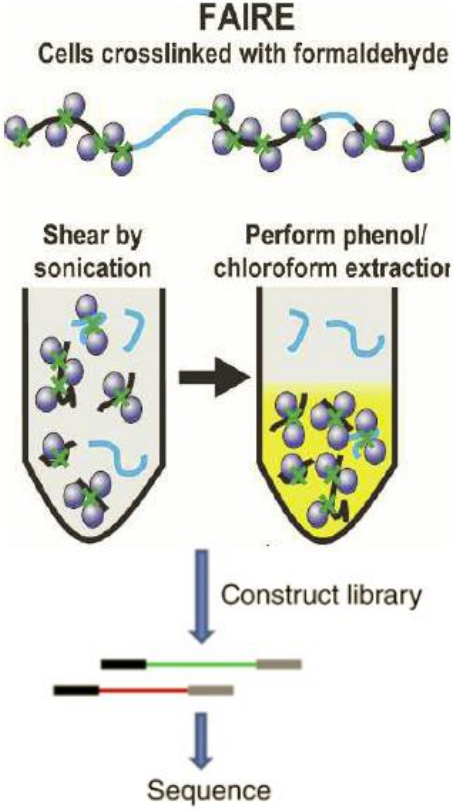


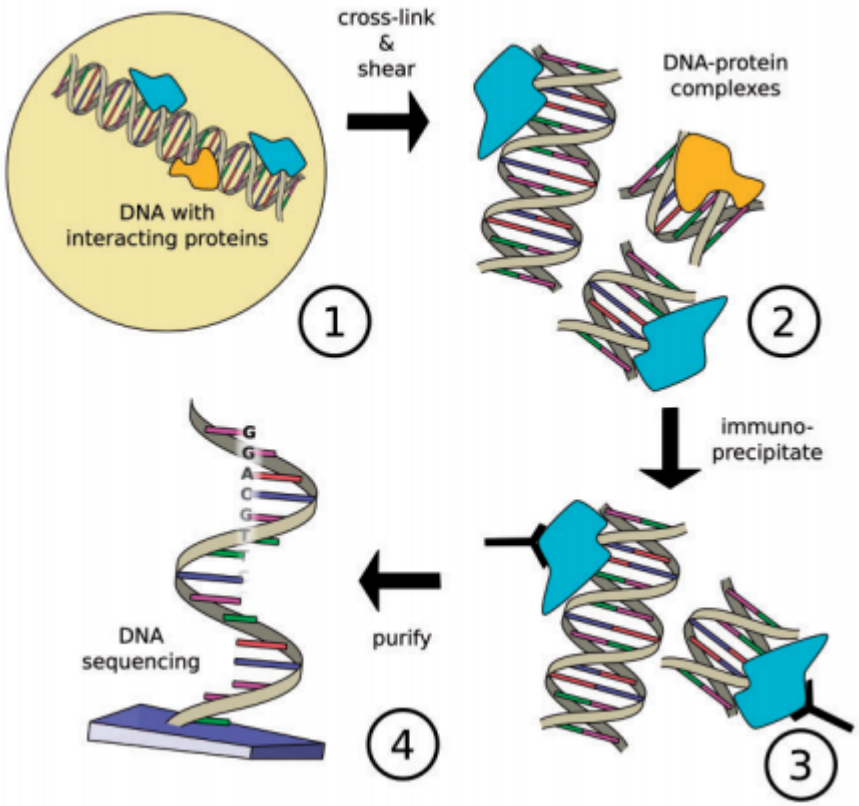
Epigenomic: Chromatin structure

Regulatory elements/ open chromatin: FAIRE-seq

Histone modifications **Protein binding sites**

ChIP-seq



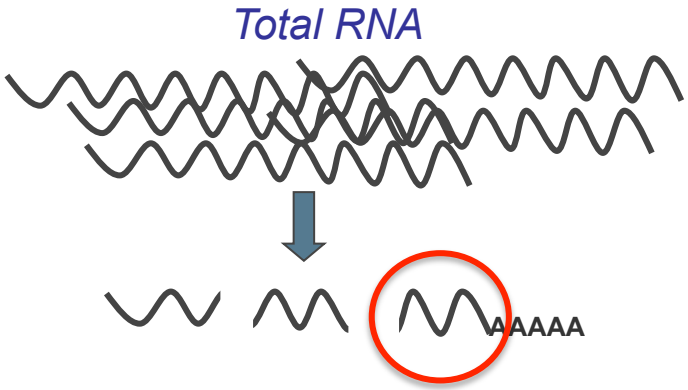


Genome Res. Jun 2007; 17(6): 877–885

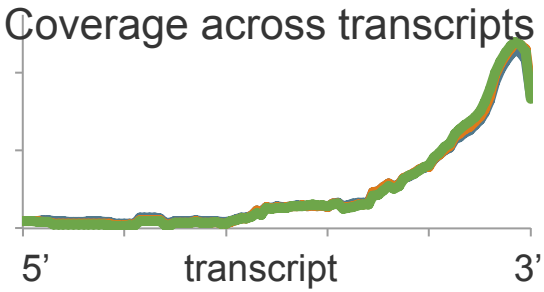
Brief Bioinform (2010) doi: 10.1093/bib/bbq068

New Technologies

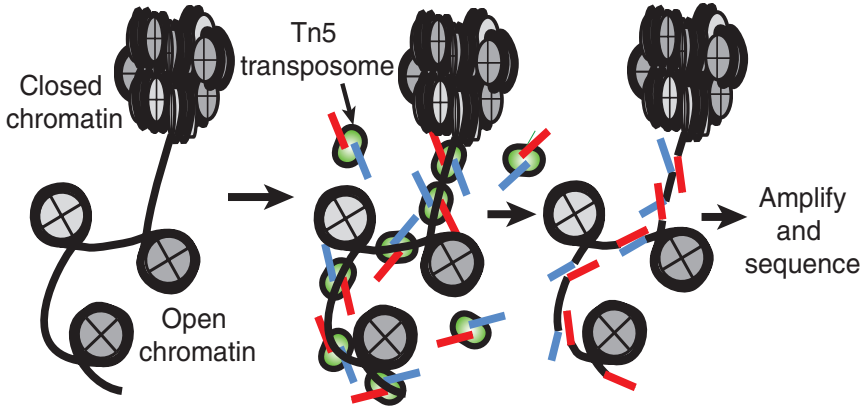
- **3’ RNA-seq**



- Gene expression study
- Reduced sequencing cost



- **ATAC-seq (Assay of Transposase Accessible Chromatin)**
  - Identification of genome wide regulatory elements



- Reduced sample requirement
- Organism dependent
  - requires customization

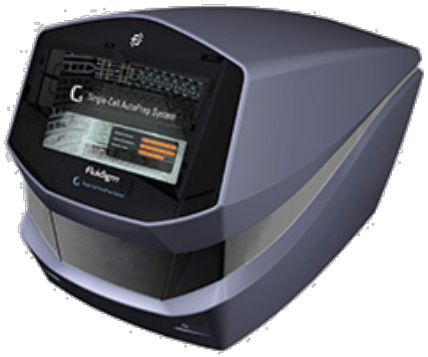
Nature Methods 10, 1213–1218 (2013)

Major User Request: Decrease Sample Amount Requirements

The need for cheaper nano-scale sample prep

Input; cost

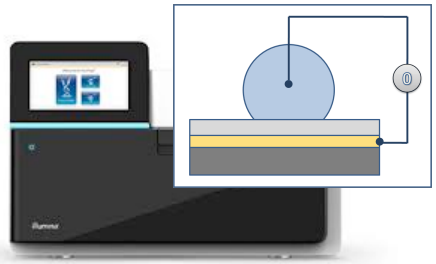
Throughput; consistency



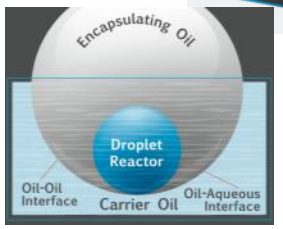
Fluidigm C1™

Echo labcyte

GenCell CLiC

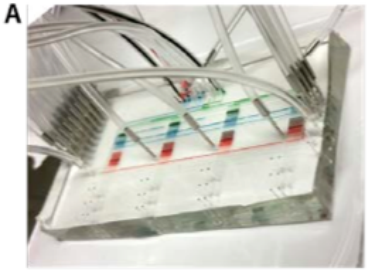


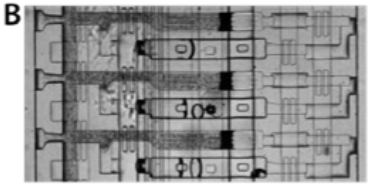
Illumina Neoprep

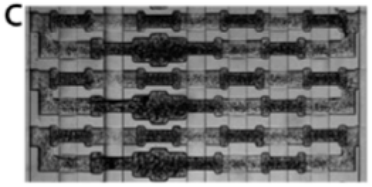


Microfluidic platform

2014-2016 ETOP
Blainey Lab MIT







Closed Commercial Microfluidic Platforms

Emerging Technologies Opportunity Program ETOP

Emerging Sequencing Technologies

- **Oxford Nanopore Technology – find out more from Juna Lee next!**